

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 27, 2003, 18:01:39 ; Search time 2272.12 Seconds  
(without alignments)  
363.693 Million cell updates/sec

Title: US-10-085-612-2  
Perfect score: 34  
Sequence: 1 cgattcttctactggtcgcagtcgagccgcg 34

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_htc.\*

9: gb\_est1.\*

10: gb\_est2.\*

11: gb\_htc.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estfun.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_pln.\*

20: em\_gss\_vrt.\*

21: em\_gss\_fun.\*

22: em\_gss\_mam.\*

23: em\_gss\_mus.\*

24: em\_gss\_pro.\*

25: em\_gss\_rod.\*

26: em\_gss\_pbg.\*

27: em\_gss\_vrt.\*

28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	24.4	71.8	802	29	BZ571326 msh2_1835
c 2	23.4	68.8	910	29	AL211728 Tetraodon
c 3	23	67.6	438	12	BM037193 fu84c01.y
c 4	23	67.6	660	9	AA739600 365 Pflg

C	5	22.6	66.5	704	29	CC011462	
6	22.2	65.3	603	9	AA803400	CC011462_PUDG231TD	
7	22.2	65.3	611	9	AA538677	AA803400 GD10776.5	
8	22	64.7	180	28	BH613513	AA538677 LD19261.5	
9	22	64.7	196	28	BH214265	BH613513 SALK_0343	
10	21.8	64.1	320	14	T60367	BH214265 SALK_0104	
11	21.8	64.1	480	9	AJ478560	T60367 BP90H09_r1	
12	21.8	64.1	509	9	AJ432574	AJ478560 AJ478560	
13	21.8	64.1	667	10	BZ730257	AJ432574 AJ432574	
14	21.8	64.1	687	12	BZ219877	BZ730257 601563625	
15	21.8	64.1	904	29	CNS04FD7	BZ219877 BJ219877	
16	21.8	64.1	946	10	BG343303	AI288196 Tetraodon	
17	21.8	64.1	972	13	BUS59835	BG343303 HVSME9000	
18	21.8	64.1	991	12	BM472533	BUS59835 AGENCOURT	
19	21.8	64.1	1043	12	BM548921	BM472533 AGENCOURT	
20	21.8	64.1	1201	29	CNS05KQ1	BM548921 AGENCOURT	
C	21	21.6	63.5	549	12	BM003553	AI341794 Tetraodon
22	21.6	63.5	681	12	BM621560	BM003553 1031111C0	
23	21.6	63.5	688	12	BM578815	BM621560 170006874	
24	21.6	63.5	730	28	AQ575468	BM578815 170006871	
25	21.4	62.9	368	28	BH253181	AQ575468 nbxd0087K	
26	21.4	62.9	422	9	AW501182	BH253181 SALK_0145	
27	21.4	62.9	485	13	BX283025	AW501182 UT-HF-BP0	
28	21.4	62.9	518	9	AI568840	BX283025 BX283025	
C	29	21.4	62.9	527	12	BH668034	AI568840 tn40e09_x
30	21.4	62.9	618	10	BF431634	BH668034 ft67a07_Y	
31	21.4	62.9	623	12	BH666970	BF431634 70a16f11_x	
32	21.4	62.9	649	10	BG424809	BH666970 ft67a07_x	
33	21.4	62.9	663	10	BG335538	BG424809 602453459	
34	21.4	62.9	683	12	B562582	BG335538 602403831	
35	21.4	62.9	687	12	BG911254	B562582 603256515	
36	21.4	62.9	713	12	B1910096	BG911254 602807519	
37	21.4	62.9	718	12	B1545634	B1910096 603067939	
38	21.4	62.9	723	10	BG717956	B1545634 603187842	
39	21.4	62.9	777	10	BF984839	BG717956 603698932	
40	21.4	62.9	798	10	BG717465	BF984839 602308094	
41	21.4	62.9	848	13	BUS94473	BG717465 602689226	
42	21.4	62.9	903	13	BUS00183	BUS94473 AGENCOURT	
43	21.4	62.9	915	14	CD358712	BUS00183 AGENCOURT	
44	21.4	62.9	929	29	CNS04LLZ	CD358712 AGENCOURT	
C	45	21.4	62.9	943	13	BU193557	AL296288 Tetraodon
46	21.4	62.9	943	13	BU193557	BU193557 AGENCOURT	

## ALIGNMENTS

RESULT 1	BZ571326	802 bp	DNA	linear	GSS 17-DEC-2002
LOCUS	msh2_1835.x1	msh	Pseudomonas aeruginosa	genomic clone msh2_1835,	
DEFINITION	genomic survey sequence.				
ACCESSION	BZ571326				
VERSION	BZ571326.1	GI:27206387			
KEYWORDS	GSS.				
SOURCE	Pseudomonas aeruginosa				
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.				
REFERENCE	1 (bases 1 to 802)				
AUTHORS	Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.				
TITLE	Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library				
JOURNAL	J. Bacteriol., (2002) In press				
COMMENT	Contact: Chris K. Raymond Genome Center University of Washington Box 352145, Seattle, WA 98105-2145, USA Tel: 2062216954 Fax: 2066857244 Email: craymond@u.washington.edu Class: shotgun.				
FEATURES	Location/Qualifiers				

## source

1. .802  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="MSH"  
 /db\_xref="taxon:287"  
 /clone\_lib="msh2\_1835"  
 /clone\_lib="msh"  
 /note="Environmental isolate. Whole genomic shotgun library."  
 BASE COUNT 139 a 218 c 164 g 228 t 53 others  
 ORIGIN

## Query Match

Best Local Similarity 71.8%; Score 24.4; DB 29; Length 802;  
 Matches 28; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

## QY

1 CGATCTTTGCTACTGCTGCAGCTGCAGCCCG 34  
 ||||| ||||| ||| ||| ||||| |||  
 348 CGATGCTTTGCTAGCGCGCGATCTGCAGCGCG 315

## RESULT 2

CNS02SD3/c  
 LOCUS  
 DEFINITION  
 Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 161J22 of library G from Tetraodon nigroviridis, genomic survey sequence.

## ACCESSION

AL2111728  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Tetraodon nigroviridis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.

## REFERENCE

1  
 Roest Crolius H., Jaillon O., Dasilva C., Bouneau L., Fisher C., Bernot A., Fizames C., Wincker P., Brottier P., Quetier F., Saurin W. and Weissenbach J.  
 Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence  
 Nat. Genet. 25 (2), 235-238 (2000)

## PUBMED

20296633  
 10835645

## REFERENCE

2  
 Roest Crolius H., Jaillon O., Dasilva C., Ozouf-Costaz C., Fizames C., Fischer C., Bouneau L., Billault A., Quetier F., Saurin W., Bernot A. and Weissenbach J.  
 Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis  
 Genome Res. 10 (7), 939-949 (2000)

## MEDLINE

20359837  
 10899143

## REFERENCE

3 (bases 1 to 910)  
 Genoscope.  
 Direct Submission  
 Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)

## TITLE

Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)

## COMMENT

This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at  
 http://www.genoscope.cns.fr/Tetraodon.

## FEATURES

Location/Qualifiers  
 1. .910  
 /organism="Tetraodon nigroviridis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:99883"  
 /clone="161J22"  
 /clone\_lib="G"  
 /note="Genoscope sequence ID : C0AG161DE11SP1-end : PUC-Ori"

## BASE COUNT

248 a 192 c 221 g 243 t 6 others  
 ORIGIN

## Query Match

Best Local Similarity 68.8%; Score 23.4; DB 29; Length 910;  
 Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

## QY

2 GATCTTTGCTACTGCTGCAGCTGCAGCCCG 34  
 ||||| ||| ||| ||||| ||||| |||  
 603 GATCTTGGCTGCAGCTGGCGCTGCAGCCCTG 571

## Db

BM037193  
 fu84c01.v1 Gong zebrafish ovary Danio rerio cDNA clone  
 IMAGE:5377536 5', similar to SW:PMW2\_HUMAN O15305 PHOSPHOMANNOMUTASE 2 ;, mRNA sequence.

## ACCESSION

BM037193  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

EST.  
 Danio rerio (zebrafish)

## REFERENCE

1 (bases 1 to 438)  
 Clark M., Johnson S.L., Lehrach H., Lee R., Li F., Marra M., Eddy S., Hillier L., Kucaba T., Martin J., Beck C., Wylie T., Underwood K., Steptoe M., Theising B., Allen M., Bowers Y., Person B., Swaller T., Gibbons M., Pape D., Harvey N., Schurk R., Ritter E., Kohn S., Shin T., Jackson Y., Cardenas M., McCann R., Waterston R. and Wilson R.  
 WashU Zebrafish EST Project 1998

## AUTHORS

Unpublished  
 Contact: Stephen L. Johnson  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: zbrafish@wustl.wustl.edu

## TITLE

The library was constructed by Dr. Z. Gong. DNA Sequencing by:

Washington University Genome Sequencing Center St. Louis. Please

contact Zhiyuan Gong for further information on this library

(National University of Singapore, Department of Biological

Sciences, Lower Kent Ridge Road, Singapore 119260).

High quality sequence stop: 323.

Location/Qualifiers

1. .438

/organism="Danio rerio"

/mol\_type="mRNA"

/db\_xref="taxon:7955"

/clone="IMAGE:5377536"

/sex="female"

/dev\_stage="4-5 month"

/lab\_host="PH10B (phage-resistant)"

/clone\_lib="Gong zebrafish ovary"

/note="Organ: ovary (pooled); Vector: pBluescript SK-"

Site\_1: XhoI; Site\_2: EcoRI; Poly A+ RNA was isolated from

the ovaries of 2 female adult zebrafish (4-5 month old).

cDNAs were made using oligo-dT primers and inserted into

lambda ZAP II vector (Stratagene) by Dr. Z. Gong, in vivo

mass-excised to pBluescript SK- following the Washington

University protocol

(http://genome.wustl.edu/est/lambda\_protocol.shtml).

Please contact Zhiyuan Gong for further information on

this library (National University of Singapore,

Department of Biological Sciences, Lower Kent Ridge Road,

Singapore 119260).

BASE COUNT 138 a 78 c 100 g 122 t

## ORIGIN

Query Match 67.6%; Score 23; DB 12; Length 438;









Search completed: August 27, 2003, 19:40:02  
Job time : 2276.12 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 27, 2003, 18:01:39 ; Search time 1202.88 Seconds  
(without alignments)  
363.693 Million cell updates/sec

Title: US-10-085-612-1  
Perfect score: 18  
Sequence: 1 gacaagggcagagacag 18

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :	EST:*
1:	em_estba.*
2:	em_esthum.*
3:	em_estin.*
4:	em_estnu.*
5:	em_estov.*
6:	em_estpl.*
7:	em_estro.*
8:	em_htc.*
9:	gb_est1.*
10:	gb_est2.*
11:	gb_htc.*
12:	gb_est3.*
13:	gb_est4.*
14:	gb_est5.*
15:	em_estfun.*
16:	em_estom.*
17:	em_gss_hum.*
18:	em_gss_inv.*
19:	em_gss_pln.*
20:	em_gss_vrt.*
21:	em_gss_fun.*
22:	em_gss_mam.*
23:	em_gss_mus.*
24:	em_gss_pro.*
25:	em_gss_rod.*
26:	em_gss_phg.*
27:	em_gss_vrl.*
28:	gb_gss1.*
29:	gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	100.0	546	28	AZ722759 RPCI-24-1
c 2	17	94.4	487	10	BG020388 dg73g11.x
c 3	17	94.4	647	28	BH305292 CH230-100
4	17	94.4	657	9	AV319601

5	17	94.4	3005	11	AK031337
6	16.4	91.1	207	10	BF925534
7	16.4	91.1	207	10	BG004130
8	16.4	91.1	245	14	CB513680
c 9	16.4	91.1	267	10	BB007362
c 10	16.4	91.1	273	10	BB008745
c 11	16.4	91.1	277	28	AZ902923
c 12	16.4	91.1	288	10	BZ006387
c 13	16.4	91.1	312	28	AZ464111
c 14	16.4	91.1	321	14	R41114
c 15	16.4	91.1	349	13	BY615518
c 16	16.4	91.1	359	14	F23252
c 17	16.4	91.1	362	13	BY135493
c 18	16.4	91.1	393	14	CB777302
c 19	16.4	91.1	413	13	BY627146
c 20	16.4	91.1	458	12	BF117896
c 21	16.4	91.1	483	10	BB761925
c 22	16.4	91.1	516	13	BQ204897
c 23	16.4	91.1	519	13	BQ495987
c 24	16.4	91.1	532	10	BG711834
c 25	16.4	91.1	580	28	AZ413382
c 26	16.4	91.1	611	28	AQ419553
c 27	16.4	91.1	620	29	BZ858741
c 28	16.4	91.1	623	29	BZ936843
c 29	16.4	91.1	640	14	CB824956
c 30	16.4	91.1	647	12	BM425824
c 31	16.4	91.1	659	29	AG046624
c 32	16.4	91.1	681	29	AG126536
c 33	16.4	91.1	699	28	BH263227
c 34	16.4	91.1	720	9	AU005711
c 35	16.4	91.1	720	28	AZ284673
c 36	16.4	91.1	771	14	CA327294
c 37	16.4	91.1	784	13	BU251593
c 38	16.4	91.1	785	13	BU128981
c 39	16.4	91.1	795	13	BU381673
c 40	16.4	91.1	824	14	CB096909
c 41	16.4	91.1	836	13	BU476726
c 42	16.4	91.1	878	13	BU108401
c 43	16.4	91.1	957	10	BG247406
c 44	16.4	91.1	1028	29	AG061505
c 45	16.4	91.1	1050	29	AG083514

ALIGNMENTS

RESULT 1  
AZ722759  
LOCUS RPCI-24-144M13-TJ RPCI-24 Mus musculus genomic clone RPCI-24-144M13  
DEFINITION 546 bp DNA linear GSS 24-JAN-2001  
Genomic survey sequence.

ACCESSION AZ722759  
VERSION AZ722759.1 GI:12466759

KEYWORDS GSS.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1. (bases 1 to 546)

AUTHORS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

Tsugaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E.,

Russell, D., de Jong, P., and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-24

Unpublished

Other GSSs: RPCI-24-144M13-TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

AK031337 Mus muscu  
BF925534 CM2-NT016  
BG004130 CM2-GN016  
CB513680 Ssa1g853  
BB007362 BB007362  
BB008745 BB008745  
AZ902923 RPCI-24-1  
BZ006387 BB006387  
AZ464111 IM0273A19  
R41114 Hk341-f adu  
BY615518 BY615518  
F23252 HSPD14635 H  
BY135493 BY135493  
CB777302 AMGNNUC.M  
BY627146 BY627146  
BF117896 BF117896  
BB761925 BE761925  
BQ204897 UI-R-EFO-  
BQ495987 EST05216  
BG711834 PGL1n-pk0  
AZ413382 IM0197K03  
AQ419553 RPCI-11-1  
BZ858741 CH240-207  
BZ936843 CH240-60H  
CB824956 R993F08.Y  
BM425824 p3f2c.pk0  
AG046624 Pan trogl  
AG126536 Pan trogl  
BH263227 CH230-121  
AU005711 AU005711  
AZ284673 RPCI-23-1  
CA327294 UI-M-FV0-  
BU251593 603403764  
BU128981 603115667  
BU381673 603861520  
CB096909 AF53-Rp1-  
BU476726 603470746  
BU108401 603110710  
BG247406 602360337  
AG061505 Pan trogl  
AG083514 Pan trogl



library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choori.org/bacpac/orderingframe.htm>). BAC end page: [http://www.tigr.org/tldb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html)  
 Plate: 144 row: M column: 13  
 Seq primer: SP6  
 Class: BAC ends.

#### FEATURES

source  
 Location/Qualifiers  
 1..546  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-24-144M13"  
 /sex="Male"  
 /cell\_type="Spleen/Brain"  
 /clone\_lib="RPCI-24"  
 /note="Vector: pTARBAC1; Site\_1: BamHI; Site\_2: BamHI; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 158 a 117 c 119 g 152 t

#### ORIGIN

Query Match 100.0%; Score 18; DB 28; Length 546;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACAGGGCAGGACAGAG 18  
 |||||  
 Db 92 GACAGGGCAGGACAGAG 109

#### RESULT 2

BG020388/c  
 LOCUS  
 DEFINITION dc73g11.x1 NICHD\_XGC\_Emb1 Xenopus laevis cDNA clone IMAGE:3402836  
 3', mRNA sequence.  
 ACCESSION BG020388  
 VERSION BG020388.1 GI:12476467  
 KEYWORDS  
 SOURCE Xenopus laevis (African clawed frog)  
 ORGANISM  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus.

#### REFERENCE

1 (bases 1 to 487)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

#### JOURNAL

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)

Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

cDNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/LLNL at: [info@image.llnl.gov](mailto:info@image.llnl.gov)

Seq primer: -400P from Gibco

High quality sequence stop: 412.

#### FEATURES

Location/Qualifiers  
 1..487  
 /organism="Xenopus laevis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8355"  
 /clone="IMAGE:3402836"  
 /tissue\_type="embryo (stage 10)"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NICHD\_XGC\_Emb1"

/note="Vector: pCMV-Sport6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert

size 1.55 kb. Constructed by Life Technologies. Note: This is a Xenopus Gene Collection (XGC) library."  
 BASE COUNT 159 a 102 c 87 g 139 t  
 ORIGIN

Query Match 94.4%; Score 17; DB 10; Length 487;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+03;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACAAGGGCAGGACAGAG 18  
 |||||  
 Db 251 ACAAGGGCAGGACAGAG 235

#### RESULT 3

BH305292/c  
 LOCUS  
 DEFINITION BH305292 TV CHORI-230 Segment 1 Rattus norvegicus genomic clone  
 CH230-100K7, genomic survey sequence.  
 ACCESSION BH305292  
 VERSION BH305292.1 GI:17217700  
 KEYWORDS  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM

Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

#### REFERENCE

1 (bases 1 to 647)  
 Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn, A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P. and Fraser, C.M.  
 Rat BAC End Sequences from Library CHORI-230 EcoRI segment

#### TITLE

#### JOURNAL

#### COMMENT

Unpublished

Contact: Shanying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: [szhao@tigr.org](mailto:szhao@tigr.org)

Clones are derived from the rat BAC library CHORI-230

(<http://www.choori.org/bacpac/rat230.htm>). For BAC library

availability, please contact Pieter de Jong ([pdejong@mail.cho.org](mailto:pdejong@mail.cho.org)).

Clones may be purchased from BACPAC Resources

(<http://www.choori.org/bacpac/orderinginformation.htm>). BAC end

page: [http://www.tigr.org/tldb/bac\\_ends/rat/bac\\_end\\_intro.html](http://www.tigr.org/tldb/bac_ends/rat/bac_end_intro.html)

Plate: 100 row: K column: 7

Seq primer: 17

Class: BAC ends.

#### FEATURES

Location/Qualifiers  
 1..647  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /strain="BN/SSNHsd/MCW"  
 /db\_xref="taxon:10116"  
 /clone="CH230-100K7"  
 /sex="Female"  
 /cell\_type="Brain"

/clone\_lib="CHORI-230 Segment 1"

/note="Vector: pTARBAC2.1; Site\_1: EcoRI; Site\_2: EcoRI;

CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by

Pieter de Jong"

BASE COUNT 157 a 143 c 193 g 154 t

#### ORIGIN

Query Match 94.4%; Score 17; DB 28; Length 647;  
 Best Local Similarity 100.0%; Pred. No. 3e+03;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACAAGGGCAGGACAGAG 18  
 |||||  
 Db 17 ACAAGGGCAGGACAGAG 1



**AUTHORS**  
 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yananaka, I., Saito, T., Okazaki, Y., Gofjohori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Glissi, C., King, B., Kochiwa, H., Quackenbush, J., Schriml, L. M., Stauble, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldairelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Cariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seva, T. C., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.

**TITLE**  
 Functional annotation of a full-length mouse cDNA collection

**JOURNAL**  
 Nature 409 (6821), 685-690 (2001)

**MEDLINE**  
 21085660

**PUBMED**  
 11217851

**REFERENCE**  
 5

**AUTHORS**  
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

**TITLE**  
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

**JOURNAL**  
 Nature 420, 563-573 (2002)

**REFERENCE**  
 6 (bases 1 to 3005)

**AUTHORS**  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirakawa, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ono, M., Ohsato, N., Okazaki, Y., Saito, R., Saiton, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

**TITLE**  
 Direct Submission

**JOURNAL**  
 Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, tel: 81-45-503-9222, Fax: 81-45-503-9216)

**COMMENT**  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ Location/Qualifiers

**FEATURES**  
 source  
 1. 3005  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM\_DB:6030408J16"  
 /db\_xref="taxon:10090"  
 /clone="6030408J16"  
 /sex="male"  
 /tissue\_type="testis"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="13 days embryo"  
 1. 3005  
 /note="fatty acid desaturase 2 (MGDI|MGDI:1930079, GB|NM\_019699, evidence: BLASTN, 100%, match=1266)"

**misc\_feature**  
 1. 3005

**BASE COUNT**  
 754 a 760 c 793 g 698 t

**ORIGIN**  
 Query Match 94.4%; Score 17; DB 11; Length 3005;  
 Best Local Similarity 100.0%; Pred. No. 4.2e+03;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**QY**  
 2 ACARAGGCGAGCAGAG 18  
 |||||

**DB**  
 2454 ACARAGGCGAGCAGAG 2470  
 |||||

**RESULT 6**  
 BF925534 207 bp mRNA linear EST 19-JAN-2001  
 CM2-NT0169-291100-525-g11 NT0169 Homo sapiens cDNA, mRNA sequence.

**DEFINITION**  
 BF925534

**ACCESSION**  
 BF925534

**VERSION**  
 BF925534.1 GI:12321936

**KEYWORDS**  
 EST.

**SOURCE**  
 Homo sapiens (human)

**ORGANISM**  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE**  
 1 (bases 1 to 207)

**AUTHORS**  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H., Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.

**TITLE**  
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

**JOURNAL**  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

**MEDLINE**  
 20202663

**PUBMED**  
 10737800

**COMMENT**  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&t2=CM2-NT0169-291100-525-g11&t3=2000-11-29&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 2  
 High quality sequence stop: 207.

**FEATURES**  
 Location/Qualifiers  
 source  
 1. 207  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"  
 /clone\_lib="NT0169"  
 /note="Organ: nervous tumor; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
 57 a 58 c 58 g 34 t

**BASE COUNT**  
 57 a 58 c 58 g 34 t

**ORIGIN**  
 Query Match 91.1%; Score 16.4; DB 10; Length 207;  
 Best Local Similarity 94.4%; Pred. No. 4.2e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

**QY**  
 1 GACARAGGCGAGCAGAG 18  
 |||||

Db 148 GACAGGGCAGCAGAG 165

RESULT 7  
BG004130  
LOCUS  
DEFINITION CM2-GN0166-201100-525-g11 GN0166 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BG004130  
VERSION BG004130.1 GI:12444997  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
PubMed 10737800  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the RAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&t2=CM2-GN0166-201100-525-g11&t3=2000-11-20&t4=1>)  
Seq primer: puc 18 forward  
High quality sequence start: 2  
High quality sequence stop: 207.  
Location/Qualifiers  
1..207  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="GN0166"  
/note="Organ: placenta\_normal; Vector: puc18; Site\_1: SmaI ; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 195,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
BASE COUNT 57 a 58 c 58 g 34 t  
ORIGIN

#### FEATURES

source

Query Match 91.18; Score 16.4; DB 10; Length 207;  
Best Local Similarity 94.4%; Pred. No. 4.2e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAGGGCAGCAGAG 18  
||||| |||||||  
DB 42 GACAGGGCAGCAGAG 59

RESULT 8  
CB513680  
LOCUS  
DEFINITION ssalrg5531327\_rev mixed\_tissue Salmo salar cDNA, mRNA sequence.  
ACCESSION CB513680  
VERSION CB513680.1 GI:29324906

Query Match 91.18; Score 16.4; DB 10; Length 207;  
Best Local Similarity 94.4%; Pred. No. 4.2e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAGGGCAGCAGAG 18  
||||| |||||||  
DB 148 GACAGGGCAGCAGAG 165

#### KEYWORDS

SOURCE

ORGANISM

EST.  
Salmo salar (Atlantic salmon)

Salmo salar

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.  
1 (bases 1 to 245)

#### REFERENCE

AUTHORS

GRASP Consortium, Davidson,W.S., Koop,B.F. and  
<http://web.uvic.ca/chr/grasp>.

A survey of Salmo salar transcripts from high complexity cDNA libraries

Unpublished

Contact: Koop BF

Centre for Biomedical Research

University of Victoria

PO Box 3020 STN CSC, Victoria BC, V8W 3N5, Canada

Tel: 250 472 4067

Fax: 250 472 4075

Email: bkoo@uvic.ca

Genome Sciences Centre, BC Cancer Agency cDNA preparation, sequencing and bioinformatics; Y Butterfield, R Kirkpatrick, J Asano, N Girn, R Guin, D Lee, S Lee, T Olson, P Pandoh, A Prabhu, D Smalls, L Spence, J Stott, S Taylor, G Yang, J Schein, S Jones and M Marra.

Location/Qualifiers

1..245

/organism="Salmo salar"

/mol\_type="mRNA"

/strain="McConnell"

/db\_xref="taxon:8030"

/clone\_lib="mixed\_tissue"

/note="vector: PCMVsp6; Library Creator: Research Genetics; Atlantic salmon tissue contributors: Carlo Biagi, Mitch Uh and Robert Devlin (DFO, Vancouver, B.C.), Simon Jones (PSS, Nanaimo, B.C.), Seaspring Hatchery (Crofton, B.C.), Rachel Roper (University of Victoria)"  
57 a 70 c 77 g 41 t

#### BASE COUNT

ORIGIN

Query Match 91.18; Score 16.4; DB 14; Length 245;  
Best Local Similarity 94.4%; Pred. No. 4.3e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAGGGCAGCAGAG 18  
||||| |||||||  
DB 42 GACAGGGCAGCAGAG 59

#### RESULT 9

BB007362/c

LOCUS

DEFINITION

musculus cDNA clone 473247D22 3', mRNA sequence.

ACCESSION

BB007362

VERSION

BB007362.1

KEYWORDS

EST.

SOURCE

ORGANISM

Mus musculus

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 267)

BB007362

RIKEN full-length enriched, 10 day neonate skin Mus

musculus cDNA clone 473247D22 3', mRNA sequence.

267 bp mRNA linear EST 22-JUN-2000

BB007362

RIKEN full-length enriched, 10 day neonate skin Mus

musculus cDNA clone 473247D22 3', mRNA sequence.

267 bp mRNA linear EST 22-JUN-2000

BB007362

RIKEN full-length enriched, 10 day neonate skin Mus

musculus cDNA clone 473247D22 3', mRNA sequence.

267 bp mRNA linear EST 22-JUN-2000

BB007362

```

TITLE          RIKEN Mouse ESTs (Konno, H., et al.)
JOURNAL        Unpublished
COMMENT        Contact: Yoshihide Hayashizaki
               Laboratory for Genome Exploration Research Group, RIKEN Genomic
               Sciences Center (GSC), Yokohama Institute
               The Institute of Physical and Chemical Research (RIKEN)
               1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
               Tel: 81-45-503-9222
               Fax: 81-45-503-9216
               Email: genome-res@sc.riken.go.jp/
               URL:http://genome.gsc.riken.go.jp/
               N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
               Thermotabilization and thermoactivation of thermostable enzymes by
               trehalose and its application for the synthesis of full length
               cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
               Itoh, M., Carninci, P., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
               Tomaru, Y., and Hayashizaki, Y.
               Automated filtration-based high-throughput plasmid preparation
               system. Genome Res. 9 (5), 463-470 (1999)
               High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
               19-44 (1999)
               Please visit our web site (http://genome.rtc.riken.go.jp) for
               further details.
FEATURES       Location/Qualifiers
SOURCE         1..267
               /organism="Mus musculus"
               /mol_type="mRNA"
               /strain="C57BL/6J"
               /db_xref="taxon:10090"
               /clone="4732477D22"
               /sex="mixed"
               /tissue_type="skin"
               /dev_stage="10 days neonate"
               /lab_host="DHI0B"
               /clone_lib="RIKEN full-length enriched, 10 day neonate
               skin"
               /note="Site.1: Sali; Site.2: BamHI; cDNA library was
               prepared and sequenced in Mouse Genome Encyclopedia
               Project of Genome Exploration Research Group in Riken
               Genomic Sciences Center and Genome Science Laboratory in
               RIKEN. Division of Experimental Animal Research in Riken
               contributed to prepare mouse tissues. 1st strand cDNA was
               primed with a primer [5'
               GAGAGAGAGAGATTCGAGCTCTTTTATTTTNN 3'], cDNA was
               prepared by using trehalose thermo-activated reverse
               transcriptase and subsequently enriched for full-length by
               cap-trapper. cDNA went through one round of normalization
               to Rot = 10.0 and subtraction to Rot = 100.0. Second
               strand cDNA was prepared with the primer adapter of
               sequence [5' GAGAGAGAGATTCGAGTAAATTAATATCCCCCCCCCC
               3']. cDNA was cloned into the XhoI and BamHI sites.
               Vector: a modified pBluescript KS(+) after bulk excision
               from Lambda FLC I"
BASE COUNT     65 a      78 c      51 g      73 t
ORIGIN
Query Match      91.1%; Score 16.4; DB 10; Length 267;
Best Local Similarity 94.4%; Pred. No. 4.4e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 1 GACAGGGCAGGACAGAG 18
    |||||
Db 130 GAGAGGGCAGGACAGAG 113

RESULT 10
BB008745/c      273 bp      mRNA      linear      EST 22-JUN-2000
LOCUS          BB008745 RIKEN full-length enriched, 10 day neonate skin Mus
DEFINITION     musculus cDNA clone 4732484105 3', mRNA sequence.

```

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BB008745  
BB008745.1 GI:8129102  
EST.  
Mus musculus (house mouse)

REFERENCE  
AUTHORS

1 (bases 1 to 273)  
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci  
P., Endo, T., Fukuda, S., Fukunishi, I., Hara, A., Hayatsu, N.,  
Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,  
Izawa, M., Kadota, K., Kawai, J., Kai, C., Kawai, J., Kikuchi, N.,  
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,  
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,  
Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata  
Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,  
Suzuki, H., Suzuki, H., Tagawa, A., Takanashi, F., Tomimaga, N., Taya  
T., Tsunoda, Y., Watanishi, A., Watanabe, S., Yamamura, T., Yamanaka, I.,  
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino  
M., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Konno, H., et al.)  
Unpublished

TITLE  
JOURNAL  
COMMENT

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp/  
URL:http://genome.gsc.riken.go.jp/  
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermotabilization and thermoactivation of thermostable enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Carninci, P., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki  
Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (http://genome.rtc.riken.go.jp) for  
further details.

FEATURES  
SOURCE

1..273  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="4732484105"  
/sex="mixed"  
/tissue\_type="skin"  
/dev\_stage="10 days neonate"  
/lab\_host="DHI0B"  
/clone\_lib="RIKEN full-length enriched, 10 day neonate  
skin"  
/note="Site.1: Sali; Site.2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGAGATTCGAGCTCTTTTATTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 10.0 and subtraction to Rot = 100.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGAGAGAGATTCGAGTAAATTAATATCCCCCCCCCC  
3']. cDNA was cloned into the XhoI and BamHI sites.  
Vector: a modified pBluescript KS(+) after bulk excision  
from Lambda FLC I"



BASE COUNT  
ORIGIN

1 GACAAGGCAGGACAGAG 18

Db 151 GAGAAGGGCAGGACAGAG 134

[illegible]

ACCESSION  
VERSION  
AZ464111.1  
GI:105222236

KEYWORDS	GSS.
SOURCE	Mus musculus (house mouse)

ORGANISM	Mus musculus
	Eukaryota; Metazoa;
	Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria;
	Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 312)
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly, M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D. Weiss,R.

**TITLE** Mouse whole genome scaffolding with paired end reads from 10kb

Journal Comment

plasmid inserts  
Unpublished  
Contact: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA

04112, 03A  
Tel: 801 585 5606

TEL: 801 585 7177  
FAX: 801 585 7177

Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)

Insert Length: 10000 · Std Error: 0.00

Plate: 0273 row: A column: 19

Seq primer: CGTTGTAACGACGCCAGT

```

class: plasmid ends

```

High quality sequence stop:

[illegible]

```
source 1. .312
```

/organism="Mus musculus"

```
/mol_type="genomic DNA"
```

```

/strain="C57BL/6J"

```

```
/db_xref="taxon:10090"
```

```
/clone="UGC1M02/3A19"
```

```

/sex="male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/note="Vector: pWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4

```

## RESULT 15

BY615518

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BY615518 349 bp mRNA linear EST 15-DEC-2002  
 BY615518 RIKEN full-length enriched, visual cortex Mus musculus  
 cDNA clone K330010005 3', mRNA sequence.

BY615518

BY615518.1 GI:26950700

EST.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 349)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
 Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H.,  
 Tagli, K., Tomaru, N., Hasegawa, Y., Nogami, A., Schonbach, C.,  
 Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A.,  
 Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S.,  
 Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbani,  
 L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest,  
 A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,  
 Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J.,  
 Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M.,  
 King, B. L., Konegaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons,  
 P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki,  
 H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G.,  
 Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D.,  
 Ramchandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring,  
 B. Z., Ringwald, M., Sanderlin, A., Schneider, C., Semple, C. A., Setou,  
 M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale,  
 R. D., Tomita, M., Verdaro, R., Wagner, L., Wahlstedt, C., Wang, Y.,  
 Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa,  
 M., Yang, I., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,  
 Carinci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura,  
 M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii,  
 Y., Itoh, M., Magawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,  
 K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,  
 E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)

12466851

COMMENT

Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane,  
 T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A.,  
 Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K.,  
 Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami,  
 M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct  
 Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Tissues were provided by Michela Fagiolini and Takao K. Hensch (  
 Laboratory for Neuronal Circuit Development Brain Science Institute  
 RIKEN 2-1 Hirotsawa, Wako-shi, Saitama 351-0198 Japan ) whose  
 assistance we gratefully acknowledge.  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
 further details.

PENTURES Location/Qualifiers  
 source 1.349

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clones="K330010005"

/tissue\_type="visual cortex"

/clone\_lib="RIKEN full-length enriched, visual cortex"

BASE COUNT 90 a 98 c 82 g 79 t

ORIGIN

Query Match 91.18: Score 16.4; DB 13: Length 349;

Best Local Similarity 94.48; Pred. No. 4.7e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAAAGGGCAGGACAGAG 18

|||||

Db 77 GACTAGGGCAGGACAGAG 94

Search completed: August 27, 2003, 19:39:58

Job time : 1205.88 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 27, 2003, 18:42:00 ; Search time 91.7308 Seconds  
(without alignments)  
450.193 Million cell updates/sec

Title: US-10-085-612-1

Perfect score: 18

Sequence: 1 gacaggaggacagacag 18

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1533700 seqs, 1147125425 residues

Total number of hits satisfying chosen parameters: 3067400

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

Published Applications\_NA.\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query % Match	Length	ID	Description
1	18	100.0	18	14	US-10-085-612-1
2	16.4	91.1	29	14	US-10-146-575-13
3	16.4	91.1	529	13	US-10-027-632-106890
4	16.4	91.1	714	13	US-10-027-632-25159
5	16.4	91.1	1530	14	US-10-198-846-10434
6	16.4	91.1	119296	14	US-10-270-336-3
7	16.4	91.1	172337	9	US-09-805-458A-3
8	16	88.9	745	13	US-10-027-632-166368
9	16	88.9	1345	14	US-10-146-575-3
10	16	88.9	1345	11	US-10-085-612-3
11	16	88.9	3881	11	US-09-948-820-20
12	15.4	85.6	19	14	US-10-146-575-44
13	15.4	85.6	299	10	US-09-783-590-3955
14	15.4	85.6	591	13	US-10-027-632-286257
15	15.4	85.6	592	13	US-10-027-632-321713
16	15.4	85.6	592	13	US-10-027-632-321714

17	15.4	85.6	618	13	US-10-027-632-105305	Sequence 105305, A
18	15.4	85.6	707	13	US-10-027-632-15402	Sequence 15402, A
19	15	83.3	654	13	US-10-027-632-275296	Sequence 275296, A
20	15	83.3	853	13	US-10-027-632-162033	Sequence 162033, A
21	15	83.3	853	13	US-10-027-632-162034	Sequence 162034, A
22	15	83.3	957	13	US-10-027-632-10403	Sequence 10403, A
23	15	83.3	957	13	US-10-027-632-10404	Sequence 10404, A
24	15	83.3	957	13	US-10-027-632-10405	Sequence 10405, A
25	14.8	82.2	18	14	US-10-085-612-5	Sequence 5, Appli
26	14.8	82.2	21	10	US-09-957-975-6	Sequence 6, Appli
27	14.8	82.2	32	14	US-10-146-575-12	Sequence 12, Appli
28	14.8	82.2	137	9	US-09-864-761-21012	Sequence 21012, A
29	14.8	82.2	349	10	US-09-867-701-2919	Sequence 2919, Ap
30	14.8	82.2	399	11	US-09-918-995-16747	Sequence 16747, A
31	14.8	82.2	429	9	US-09-864-761-4559	Sequence 4259, Ap
32	14.8	82.2	430	13	US-10-027-632-271248	Sequence 271248, A
33	14.8	82.2	503	13	US-10-027-632-311598	Sequence 311598, A
34	14.8	82.2	509	13	US-10-027-632-682234	Sequence 68234, A
35	14.8	82.2	542	10	US-09-867-701-5985	Sequence 5985, Ap
36	14.8	82.2	590	13	US-10-027-632-59486	Sequence 59486, A
37	14.8	82.2	590	13	US-10-027-632-59487	Sequence 59487, A
38	14.8	82.2	590	13	US-10-027-632-59488	Sequence 59488, A
39	14.8	82.2	597	9	US-09-864-761-7390	Sequence 7390, Ap
40	14.8	82.2	598	13	US-10-027-632-227843	Sequence 227843, A
41	14.8	82.2	625	13	US-10-027-632-38017	Sequence 38017, A
42	14.8	82.2	637	13	US-10-027-632-86427	Sequence 86427, A
43	14.8	82.2	637	13	US-10-027-632-316370	Sequence 316370, A
44	14.8	82.2	653	13	US-10-027-632-70764	Sequence 70764, A
45	14.8	82.2	720	13	US-10-027-632-72506	Sequence 72506, A

#### ALIGNMENTS

#### RESULT 1

US-10-085-612-1  
; Sequence 1, Application US/10085612  
; Publication No. US20030096251A1  
; GENERAL INFORMATION:  
; APPLICANT: Guida, Marco  
; APPLICANT: Hall, Jeff  
; APPLICANT: Petros, William  
; APPLICANT: Vredenburg, James  
; APPLICANT: Vredenburg, James  
; APPLICANT: Colvin, Oliver  
; APPLICANT: Marks, Jeffrey  
; TITLE OF INVENTION: Methods for Evaluating the Ability to Metabolize Pharmaceutica  
; FILE REFERENCE: 4389-5-C1  
; CURRENT APPLICATION NUMBER: US/10/085,612  
; CURRENT FILING DATE: 2002-02-26  
; PRIOR FILING DATE: 1998-08-31  
; PRIOR FILING DATE: 2001-02-26  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 18  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-085-612-1

Query Match 100.08; Score 18; DB 14; Length 18;  
Best Local Similarity 100.08; Pred. No. 18;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACAGGGCAGCAGACAG 18

Db 1 GACAGGGCAGCAGACAG 18

#### RESULT 2

US-10-146-575-13/c

```

; Sequence 13, Application US/10146575
; Publication No. US20030059800A1
; GENERAL INFORMATION:
; APPLICANT: Lichter, Jay
; APPLICANT: Guido, Marco
; TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4
; FILE REFERENCE: SEQ-12P
; CURRENT APPLICATION NUMBER: US/10/146,575
; CURRENT FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: US/09/144,367
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 29
; TYPE: DNA
; ORGANISM: H. sapiens
US-10-146-575-13

```

```

Query Match          91.1%; Score 16.4; DB 14; Length 29;
Best Local Similarity 94.4%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 GACAAGGGCAGGACAGAG 18
        |||||
Db       24 GACAAGGGCAGGACAGAG 7

```

```

RESULT 3
US-10-027-632-106890/c
; Sequence 106890, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 10827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 106890
; LENGTH: 529
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-106890

```

```

Query Match          91.1%; Score 16.4; DB 13; Length 529;
Best Local Similarity 94.4%; Pred. No. 76;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 GACAAGGGCAGGACAGAG 18
        |||||
Db       170 GTCAGGGCAGGACAGAG 153

```

```

RESULT 4
US-10-027-632-25159
; Sequence 25159, Application US/10027632
; GENERAL INFORMATION:

```

```

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; GENERAL INFORMATION: Polymorphisms in the Human Genome
; FILE REFERENCE: 10827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25159
; LENGTH: 714
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-25159

```

```

Query Match          91.1%; Score 16.4; DB 13; Length 714;
Best Local Similarity 94.4%; Pred. No. 74;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 GACAAGGGCAGGACAGAG 18
        |||||
Db       361 GTCAGGGCAGGACAGAG 378

```

```

RESULT 5
US-10-198-846-10434
; Sequence 10434, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Yongyao
; APPLICANT: Lillie, James
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10434
; LENGTH: 1530
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 897..1528, 1529..1530
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-10434

```

```

Query Match          91.1%; Score 16.4; DB 14; Length 1530;
Best Local Similarity 94.4%; Pred. No. 69;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 GACAAGGGCAGGACAGAG 18
        |||||
Db       997 GACGAGGGCAGGACAGAG 1014

```

```
RESULT 6
US-10-270-336-3
; Sequence 3, Application US/10270336
; Publication No. US20030074678A1
; GENERAL INFORMATION:
; APPLICANT: ZHU, Shiaoqing et al.
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: C1001146CON
; CURRENT APPLICATION NUMBER: US/10/270,336
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 60/270,873
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 119596
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(119596)
; OTHER INFORMATION: n = A,T,C or G
US-10-270-336-3

Query Match          91.1%; Score 16.4; DB 14; Length 119596;
Best Local Similarity 94.4%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAAGGGCAGGACAGAG 18
Db 63483 GACAAGGGCAGGACAGAG 63500

RESULT 7
US-09-805-458A-3
; Sequence 3, Application US/09805458A
; Patent No. US20020042100A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al
; TITLE OF INVENTION: ISOLATED HUMAN ION CHANNEL PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN ION CHANNEL PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: C1000722
; CURRENT APPLICATION NUMBER: US/09/805,458A
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 172637
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(172637)
; OTHER INFORMATION: n = A,T,C or G
US-09-805-458A-3

Query Match          91.1%; Score 16.4; DB 9; Length 172637;
Best Local Similarity 94.4%; Pred. No. 44;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAAGGGCAGGACAGAG 18
Db 95503 GACAAGGGCAGGACAGAG 95520

RESULT 8
US-10-027-632-166368/c
; Sequence 166368, Application US/10027632
```

```
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 166368
; LENGTH: 745
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-166368

Query Match          88.9%; Score 16; DB 13; Length 745;
Best Local Similarity 88.9%; Pred. No. 1.2e+02;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAAGGGCAGGACAGAG 18
Db 46 GTCAAGGGCAGGACAGAG 29

RESULT 9
US-10-146-575-3
; Sequence 3, Application US/10146575
; Publication No. US20030059800A1
; GENERAL INFORMATION:
; APPLICANT: Lichter, Jay
; APPLICANT: Guido, Marco
; TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4
; FILE REFERENCE: SEQ-12P
; CURRENT APPLICATION NUMBER: US/10/146,575
; CURRENT FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: US/09/144,367
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1345
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: Other
; LOCATION: (0)...(0)
US-10-146-575-3

Query Match          88.9%; Score 16; DB 14; Length 1345;
Best Local Similarity 88.9%; Pred. No. 1.1e+02;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAAGGGCAGGACAGAG 18
Db 806 GACAAGGGCAGGACAGAG 823

RESULT 10
US-10-085-612-3
```

```
; Sequence 3, Application US/10085612
; Publication No. US20030096251A1
; GENERAL INFORMATION:
; APPLICANT: Guida, Marco
; APPLICANT: Hall, Jeff
; APPLICANT: Petros, William
; APPLICANT: Vredenburg, James
; APPLICANT: Colvin, Oliver
; APPLICANT: Marks, Jeffrey
; TITLE OF INVENTION: Methods for Evaluating the Ability to Metabolize Pharmaceuticals
; FILE REFERENCE: 4389-5-C1
; CURRENT APPLICATION NUMBER: US/10/085,612
; PRIOR FILING DATE: 2002-02-26
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: 60/271,630
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1345
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-085-612-3
```

```
Query Match      88.9%; Score 16; DB 14; Length 1345;
Best Local Similarity 88.9%; Pred. No. 1.1e+02;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1  GACAGGGCAGCAGACAG 18
        |||||
Db      806  GACAGGGCAGCAGACAG 823
```

```
RESULT 11
US-09-948-820-20/c
; Sequence 20, Application US/09948820
; Publication No. US20030050460A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; FILE REFERENCE: P2034P1
; CURRENT APPLICATION NUMBER: US/09/948,820
; CURRENT FILING DATE: 2001-09-10
; PRIOR FILING DATE: 2000-05-05
; PRIOR FILING DATE: 1999-11-09
; PRIOR FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 3881
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-948-820-20
```

```
Query Match      88.9%; Score 16; DB 11; Length 3881;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      3  CAAGGCGCAGCAGACAG 18
        |||||
Db      2592  CAAGGCGCAGCAGACAG 2577
```

```
RESULT 12
US-10-146-575-44
; Sequence 44, Application US/10146575
; Publication No. US20030059800A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Lichter, Jay
; APPLICANT: Guido, Marco
; TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4
; FILE REFERENCE: SEQ-12P
; CURRENT APPLICATION NUMBER: US/10/146,575
; CURRENT FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: US/09/144,367
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 44
; LENGTH: 19
; TYPE: DNA
; ORGANISM: H. sapiens
US-10-146-575-44
```

```
Query Match      85.6%; Score 15.4; DB 14; Length 19;
Best Local Similarity 94.1%; Pred. No. 3.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      2  ACAAGGCGCAGCAGACAG 18
        |||||
Db      1  ACAAGGCGCAGCAGACAG 17
```

```
RESULT 13
US-09-783-590-3955
; Sequence 3955, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2c1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3955
; LENGTH: 299
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (8)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (15)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (23)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (45)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (97)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (123)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (131)
```

```

; OTHER INFORMATION: n equals a.t.g, or c
; NAME/KEY: misc feature
; LOCATION: (132)
; OTHER INFORMATION: n equals a.t.g, or c
; NAME/KEY: misc feature
; LOCATION: (204)
; OTHER INFORMATION: n equals a.t.g, or c
; NAME/KEY: misc feature
; LOCATION: (214)
; OTHER INFORMATION: n equals a.t.g, or c
; NAME/KEY: misc feature
; LOCATION: (267)
; OTHER INFORMATION: n equals a.t.g, or c
; NAME/KEY: misc feature
; LOCATION: (271)
; OTHER INFORMATION: n equals a.t.g, or c
; NAME/KEY: misc feature
; LOCATION: (294)
; OTHER INFORMATION: n equals a.t.g, or c
US-09-783-590-3955

```

```

Query Match      85.6%; Score 15.4; DB 10; Length 299;
Best Local Similarity 88.9%; Pred. No. 2.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      1 GACAAGGCGCAGGACAGAG 18
         || ||| ||||| ||||| |||
Db      39 GAAGAGCGCAGGACAGAG 56

```

## RESULT 14

```

US-10-027-632-286257/c
; Sequence 286257, Application US/10027632
; GENERAL INFORMATION:

```

```

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 286257
; LENGTH: 591
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-286257

```

```

Query Match      85.6%; Score 15.4; DB 13; Length 591;
Best Local Similarity 94.1%; Pred. No. 2.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 GACAAGGCGCAGGACAGAG 17
         || ||||| ||||| |||
Db      303 GAAGAGCGCAGGACAGAG 287

```

## RESULT 15

```

US-10-027-632-321713

```

```

; Sequence 321713, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 321713
; LENGTH: 592
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-321713

```

```

Query Match      85.6%; Score 15.4; DB 13; Length 592;
Best Local Similarity 94.1%; Pred. No. 2.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 GACAAGGCGCAGGACAGAG 17
         ||||| ||||| |||
Db      244 GACAAGAGCAGGACAGAG 260

```

```

Search completed: August 27, 2003, 21:14:03
Job time : 92.7308 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 27, 2003, 18:00:44 ; Search time 1013.46 Seconds  
(without alignments)  
1372.452 Million cell updates/sec

Title: US-10-085-612-2

Perfect score: 34  
Sequence: 1 cgaattcttgcactgcagctgcagcccc 34

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters:	5777422
--	---------

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gp.ba.\*
- 2: gp.hcg.\*
- 3: gp.in.\*
- 4: gp.in.\*
- 5: gp.ov.\*
- 6: gp.pat.\*
- 7: gp.ph.\*
- 8: gp.pl.\*
- 9: gp.pr.\*
- 10: gp.ro.\*
- 11: gp.sts.\*
- 12: gp.sy.\*
- 13: gb.un.\*
- 14: gb.yi.\*
- 15: en.ba.\*
- 16: en.fun.\*
- 17: en.hum.\*
- 18: en.in.\*
- 19: en.mu.\*
- 20: en.om.\*
- 21: en.or.\*
- 22: en.ov.\*
- 23: en.pat.\*
- 24: en.ph.\*
- 25: en.pl.\*
- 26: en.ro.\*
- 27: en.sts.\*
- 28: en.un.\*
- 29: en.yi.\*
- 30: en.hcg.hum.\*
- 31: en.hcg.inv.\*
- 32: en.hcg.other.\*
- 33: en.hcg.mus.\*
- 34: en.hcg.pln.\*
- 35: en.hcg.rod.\*
- 36: en.hcg.mam.\*
- 37: en.hcg.vrt.\*
- 38: en.sy.\*
- 39: en.hcg.hum.\*
- 40: en.hcg.mus.\*
- 41: en.hcg.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	34	100.0	830	6	AX472122	Sequence
2	34	100.0	1612	9	AF325929	Homo sapi
3	34	100.0	123778	9	AC069994	Homo sapi
4	33	97.1	442	9	AF315320S4	Homo sapi
5	33	97.1	830	6	AX472123	Sequence
6	33	97.1	830	6	AX472124	Sequence
7	33	97.1	830	6	AX472216	Sequence
8	33	97.1	830	6	AX472217	Sequence
9	33	97.1	830	6	AX472218	Sequence
10	33	97.1	98960	6	AX706964	Sequence
11	33	97.1	95960	6	AX707894	Sequence
12	33	97.1	174832	9	AF280107	Homo sapi
13	31.4	92.4	830	6	AX472125	Sequence
14	31.4	92.4	830	6	AX472244	Sequence
15	30.8	90.6	174004	2	AC141417	Papio anu
16	27.6	81.2	621	6	AX472134	Sequence
17	27.6	81.2	624	6	AX472132	Sequence
18	27.6	81.2	624	6	AX472133	Sequence
19	27.6	81.2	624	6	AX472235	Sequence
20	27.6	81.2	624	6	AX472236	Sequence
21	27.6	81.2	177531	6	AX706962	Sequence
22	27.6	81.2	177531	9	AX707892	Sequence
23	27.6	81.2	177531	9	AC005020	Homo sapi
24	23	67.6	110000	2	AC095225_0	Rattus nor
25	23	67.6	157888	2	AC104927	Rattus nor
26	23	67.6	171724	2	AC129544	Mus muscu
27	23	67.6	251207	2	AC106955	Rattus nor
28	23	67.6	255495	2	AC128973	Rattus nor
29	23	67.6	271429	2	AC129086	Rattus nor
30	23	67.6	280198	2	AC195813	Rattus nor
31	22.8	67.1	1012	9	AF181861	Homo sapi
32	22.8	67.1	1013	9	AF181105	Homo sapi
33	22.8	67.1	1345	6	AF142139	Sequence
34	22.8	67.1	1345	6	AF142140	Sequence
35	22.8	67.1	1345	6	AR222893	Sequence
36	22.8	67.1	1345	6	AX421253	Sequence
37	22.8	67.1	1345	9	HUMCPY3A4	Homo sapi
38	22.8	67.1	111866	9	AF329900	Homo sapi
39	22.8	67.1	11374	9	AF185589	Homo sapi
40	22.4	65.9	625	11	BV031220	S212P604
41	22.4	65.9	89212	9	AC016679	Homo sapi
42	22.4	65.9	138902	9	HSAB13H19	Human DN
43	22.4	65.9	154125	2	AC073227	Homo sapi
44	22.4	65.9	168587	9	AC090810	Homo sapi
45	22.4	65.9	177434	2	AC067832	Homo sapi

## ALIGNMENTS

[illegible]

JOURNAL Patent: WO 02053775-A 113 11-JUL-2002;

EPIDAUROS BIOTECHNOLOGIE AG (DE)

## FEATURES

source  
Location/Qualifiers  
1. .830  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

BASE COUNT 229 a 176 c 239 g 186 t  
ORIGIN

Query Match 100.0%; Score 34; DB 6; Length 830;

Best Local Similarity 100.0%; Pred. No. 0.0012;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATTCTTGTCTACTGGCTGCAGCTGCAGCCCG 34

Db 549 CGATTCTTGTCTACTGGCTGCAGCTGCAGCCCG 582

## RESULT 2

AF325929

LOCUS

DEFINITION Homo sapiens cytochrome P450 (CYP3AP1) pseudogene, CYP3AP1\*1

ACCESSION AF325929

VERSION AF325929.1

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

REFERENCE

1 (bases 1 to 1612)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.

Kuehl, P., Zhang, J., Lin, Y., Lamba, J., Assem, M., Schuetz, J.,

Watkins, P.B., Daly, A., Wrighton, S.A., Hall, S.D., Maurel, P.,

Relling, M., Brimer, C., Yasuda, K., Venkataraman, R., Storm, S.,

Thummel, K., Boguski, M.S., and Schuetz, E.

Sequence diversity in CYP3A promoters and characterization of the

genetic basis of polymorphic CYP3A5 expression

Nat. Genet. 27 (4), 383-391 (2001)

21175742

PUBMED 11279519

REFERENCE

2 (bases 1 to 1612)

Kuehl, P.M., Zhang, J., Lin, Y., Watkins, P., Maurel, P., Daly, A.,

Wrighton, S., Hall, S., Relling, M., Schuetz, J., Brimer, C., Yasuda, K.,

Storm, S., Thummel, K., Schuetz, E., and Boguski, M.

Direct Submission

Submitted (05-DEC-2000) University of Maryland at Baltimore,

Baltimore, MD 21201, USA

Location/Qualifiers

1. .1612

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

1. .>1612

/gene="CYP3AP1"

/pseudo

/allele="CYP3AP1\*1"

1. .1436

/gene="CYP3AP1"

/pseudo

1374. .1378

/note="2 base mismatch and 2 base insertion compared to

GenBank Accession Number S74700"

/evidence=experimental

1340. .>1612

/gene="CYP3AP1"

/note="cytochrome P450"

/pseudo

/codon\_start=1

494 a 334 c 388 g 396 t

BASE COUNT 494 a 334 c 388 g 396 t

ORIGIN

Query Match 100.0%; Score 34; DB 9; Length 1612;

Best Local Similarity 100.0%; Pred. No. 0.0012;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATTCTTGTCTACTGGCTGCAGCTGCAGCCCG 34

Db 1360 CGATTCTTGTCTACTGGCTGCAGCTGCAGCCCG 1393

## RESULT 3

AC069294

LOCUS

DEFINITION Homo sapiens BAC clone RP11-757A13 from 7, complete sequence.

ACCESSION AC069294

VERSION AC069294.5

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

REFERENCE

1 (bases 1 to 123778)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.

Sulston, J.E. and Waterston, R.

Toward a complete human genome sequence

Genome Res. 8 (11), 1097-1108 (1998)

99063792

PUBMED 9847074

REFERENCE

2 (bases 1 to 123778)

Bielicki, L. and Cotton, M.

The sequence of Homo sapiens BAC clone RP11-757A13

Unpublished

REFERENCE

3 (bases 1 to 123778)

Waterston, R.H.

Direct Submission

Submitted (24-MAY-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

REFERENCE

4 (bases 1 to 123778)

Waterston, R.H.

Direct Submission

Submitted (23-FEB-2001) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

REFERENCE

5 (bases 1 to 123778)

Waterston, R.

Direct Submission

Submitted (09-MAY-2001) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Feb 23, 2001 this sequence version replaced gi:9858450.

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu/gsc

Contact: sapiens@wustl.edu

Summary Statistics

Center project name: H\_NH0757A13

NOTICE: This sequence may not represent the entire insert of this

clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap

between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:

all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by sequence

from more than one subclone; and the assembly was confirmed by

restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and

sequencing collaboration between the NHGRI Chromosome 7 Mapping

Project (Eric D. Green, Director), John D. McPherson in the

Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/OTB/CHR7>, send mailto:egreen@hgri.nih.gov, or see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

## VECTOR: pBACE3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is GS1-259H13, 200 bp overlap; the clone sequenced to the left is CPD-3244O18, 200 bp overlap. Actual start of this clone is at base position 59660 of CPD-3244O18; actual end is at base position 25744 of GS1-259H13.

The sequence from base position 113692 to 114784 is derived from a single plasmid subclone. Assembly in this region is supported by HindIII digest information.

## FEATURES

## Location/Qualifiers

## source

1...123778

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="7"

/map="7"

/clone="RP11-757A13"

/clone\_lib="RPCI-11"

1...337

/rpt\_family="MaLR"

353...657

/rpt\_family="Alu"

814...928

/rpt\_family="L2"

1684...1900

/rpt\_family="Alu"

2040...2334

/rpt\_family="Alu"

2453...2756

/rpt\_family="Alu"

3874...4033

/rpt\_family="MER1\_type"

10188...10513

/note="similar to EST AI377273 (NID:g4187126) te65b04.xl"

10244...10541

/note="similar to EST BF086380 (NID:g10892090)"

10245...10562

/note="similar to EST AW977693 (NID:g8169072)"

10255...10556

/note="similar to EST BE218528 (NID:g9905846) hv40h11.xl"

10257...10565

/note="similar to EST BF093582 (NID:g10899292)"

10319...10568

/note="similar to EST AA934962 (NID:g3092174) op32q05.sl"

10569...10682

/rpt\_family="Alu"

10776...10934

/rpt\_family="Alu"

10936...11009

/note="similar to EST BE218528 (NID:g8905846) hv40h11.xl"

10942...11021

/note="similar to EST AW977693 (NID:g8169072)"

11179...11348

/rpt\_family="ERV1"

11537...11822

/rpt\_family="ERV1"

11814...12038

repeat\_region /note="similar to EST AW890805 (NID:g8055010)" 12025...12094  
 repeat\_region /rpt\_family="ERV1" 12095...12505  
 repeat\_region /rpt\_family="ERV1" 12509...14044  
 repeat\_region /rpt\_family="L1" 14045...14130  
 repeat\_region /rpt\_family="Alu" 14137...15361  
 repeat\_region /rpt\_family="L1" 15351...15634  
 misc\_feature /note="similar to EST AI140604 (NID:g3648061) qe05f02.xl" 15354...15634  
 misc\_feature /note="similar to EST AA811274 (NID:g2880885) ob68e03.sl" 15357...15634  
 misc\_feature /note="similar to EST AA858281 (NID:g2946583) ob13a02.sl" 15358...15634  
 misc\_feature /note="similar to EST AA935054 (NID:g3091761) oo68d11.sl" 15373...15634  
 repeat\_region /note="similar to EST BE682487 (NID:g10068399)" 19623...19743  
 repeat\_region /rpt\_family="L1" 19846...20329  
 repeat\_region /rpt\_family="L1" 20331...20499  
 repeat\_region /rpt\_family="L1" 20499...21923  
 repeat\_region /rpt\_family="L1" 22194...23073  
 repeat\_region /rpt\_family="L1" 23074...23663  
 repeat\_region /rpt\_family="L1" 23524...23650  
 misc\_feature /note="similar to EST AA570297 (NID:g2344277) nm07h03.sl" 24661...24717  
 repeat\_region /rpt\_family="MIR" 24820...25131  
 repeat\_region /rpt\_family="Alu" 25243...25978  
 repeat\_region /rpt\_family="L1" 26005...26043  
 repeat\_region /rpt\_family="L1" 26402...26483  
 repeat\_region /rpt\_family="L1" 26485...26797  
 repeat\_region /rpt\_family="Alu" 27787...27913  
 repeat\_region /rpt\_family="L1" 27936...28269  
 repeat\_region /rpt\_family="L1" 28292...28354  
 repeat\_region /rpt\_family="MaLR" 28384...28690  
 repeat\_region /rpt\_family="Alu" 28815...28939  
 repeat\_region /rpt\_family="L2" 29027...29294  
 repeat\_region /rpt\_family="L2" 30023...30152  
 repeat\_region /rpt\_family="MaLR" 30320...30377  
 repeat\_region /rpt\_family="MIR" 30396...30596

Query Match 100.0%; Score 34; DB 9; Length 123778;

Best Local Similarity 100.0%; Pred No. 0.001;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATTCTTGTCTACTGGCTGACGTGACGCCCG 34

|||||

DB 119173 CGATTCTTGTCTACTGGCTGACGTGACGCCCG 119206



```

RESULT 4
AF31532054          442 bp   DNA       linear   PRI 07-FEB-2001
LOCUS               Homo sapiens CYP3A2 pseudogene, partial sequence.
DEFINITION          AF315323
ACCESSION           AF315323
VERSION             AF315323.1 GI:12082805
KEYWORDS
SEGMENT
SOURCE              4 of 5
                   Homo sapiens (human)
ORGANISM            Homo sapiens
                   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE           1 (bases 1 to 442)
AUTHORS             Finta,C. and Zaphiropoulos,P.G.
TITLE               The human cytochrome P450 3A locus. Gene evolution by capture of
                   downstream exons
JOURNAL             Gene 260 (1-2), 13-23 (2000)
MEDLINE            2057884
PUBMED             11137287
REFERENCE           2 (bases 1 to 442)
AUTHORS             Finta,C. and Zaphiropoulos,P.G.
TITLE               Direct Submission
JOURNAL             Submitted (23-OCT-2000) Department of Biosciences, Karolinska
                   Institute, NOVUM, Huddinge 14157, Sweden
FEATURES             Location/Qualifiers
                     source
                     1..442
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21-q22.1"
                     /clone="CITE BAC 128D24"
                     /gene="CYP3A2"
                     /note="exon 1-like sequence; mono-oxygenase"
                     /pseudo
BASE COUNT          112 a 106 c 111 g 113 t
ORIGIN
misc_feature
Query Match          97.1%; Score 33; DB 9; Length 442;
Best Local Similarity 100.0%; Pred. No. 0.0035;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GATTCCTTGCTACTGCTGCAGCTGCAGCCCG 34
|||||
DB 67 GATTCCTTGCTACTGCTGCAGCTGCAGCCCG 99

RESULT 5
AX472123
LOCUS               AX472123          830 bp   DNA       linear   PAT 09-AUG-2002
DEFINITION          Sequence 114 from Patent WO02053775.
ACCESSION           AX472123
VERSION             AX472123.1 GI:22207164
KEYWORDS
SOURCE              Homo sapiens (human)
ORGANISM            Homo sapiens
                   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE           1
AUTHORS             Hustert,E., Haberl,M. and Wojnowski,L.
TITLE               Identification of the genetic determinants of the polymorphic
                   CYP3A5 expression
JOURNAL             Patent: WO 02053775-A 114 11-JUL-2002;
                   EPIDAURUS BIOTECHNOLOGIE AG (DE)
FEATURES             Location/Qualifiers
                     source
                     1..830
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT          231 a 176 c 236 g 187 t
ORIGIN
Query Match          97.1%; Score 33; DB 9; Length 442;
Best Local Similarity 100.0%; Pred. No. 0.0035;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GATTCCTTGCTACTGCTGCAGCTGCAGCCCG 34
|||||
DB 67 GATTCCTTGCTACTGCTGCAGCTGCAGCCCG 99

RESULT 6
AX472124
LOCUS               AX472124          830 bp   DNA       linear   PAT 09-AUG-2002
DEFINITION          Sequence 115 from Patent WO02053775.
ACCESSION           AX472124
VERSION             AX472124.1 GI:22207165
KEYWORDS
SOURCE              Homo sapiens (human)
ORGANISM            Homo sapiens
                   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE           1
AUTHORS             Hustert,E., Haberl,M. and Wojnowski,L.
TITLE               Identification of the genetic determinants of the polymorphic
                   CYP3A5 expression
JOURNAL             Patent: WO 02053775-A 115 11-JUL-2002;
                   EPIDAURUS BIOTECHNOLOGIE AG (DE)
FEATURES             Location/Qualifiers
                     source
                     1..830
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT          231 a 176 c 236 g 187 t
ORIGIN
Query Match          97.1%; Score 33; DB 6; Length 830;
Best Local Similarity 100.0%; Pred. No. 0.0034;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATTCCTTGCTACTGCTGCAGCTGCAGCCCG 33
|||||
DB 549 CGATTCCTTGCTACTGCTGCAGCTGCAGCCCG 581

RESULT 7
AX472216
LOCUS               AX472216          830 bp   DNA       linear   PAT 09-AUG-2002
DEFINITION          Sequence 207 from Patent WO02053775.
ACCESSION           AX472216
VERSION             AX472216.1 GI:22207253
KEYWORDS
SOURCE              Homo sapiens (human)
ORGANISM            Homo sapiens
                   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE           1
AUTHORS             Hustert,E., Haberl,M. and Wojnowski,L.
TITLE               Identification of the genetic determinants of the polymorphic
                   CYP3A5 expression
JOURNAL             Patent: WO 02053775-A 207 11-JUL-2002;
                   EPIDAURUS BIOTECHNOLOGIE AG (DE)
FEATURES             Location/Qualifiers
                     source
                     1..830
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT          230 a 177 c 237 g 186 t
ORIGIN
Query Match          97.1%; Score 33; DB 6; Length 830;
Best Local Similarity 100.0%; Pred. No. 0.0034;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATTCCTTGCTACTGCTGCAGCTGCAGCCCG 33
|||||
DB 549 CGATTCCTTGCTACTGCTGCAGCTGCAGCCCG 581

```

QY 1 CGATTCTTGTCTACTGGCTGCAGCTGCAGCCCG 33  
 |||||||  
 Db 549 CGATTCTTGTCTACTGGCTGCAGCTGCAGCCCG 581

RESULT 8  
 AX472217  
 LOCUS AX472217 830 bp DNA PAT 09-AUG-2002  
 DEFINITION Sequence 208 from Patent WO02053775.  
 ACCESSION AX472217  
 VERSION AX472217.1 GI:22207254  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 AUTHORS Hustert,E., Haberl,M. and Wojnowski,L.  
 TITLE Identification of the genetic determinants of the polymorphic  
 cyp3a5 expression  
 JOURNAL Patent: WO 02053775-A 208 11-JUL-2002;  
 EPIDAURUS BIOTECHNOLOGIE AG (DE)  
 FEATURES  
 source  
 1. .830  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 230 a 175 c 237 g 188 t  
 ORIGIN

Query Match 97.1%; Score 33; DB 6; Length 830;  
 Best Local Similarity 100.0%; Pred. No. 0.0034;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATTCTTGTCTACTGGCTGCAGCTGCAGCCCG 33  
 |||||||  
 Db 549 CGATTCTTGTCTACTGGCTGCAGCTGCAGCCCG 581

RESULT 9  
 AX472218  
 LOCUS AX472218 830 bp DNA PAT 09-AUG-2002  
 DEFINITION Sequence 209 from Patent WO02053775.  
 ACCESSION AX472218  
 VERSION AX472218.1 GI:22207255  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 AUTHORS Hustert,E., Haberl,M. and Wojnowski,L.  
 TITLE Identification of the genetic determinants of the polymorphic  
 cyp3a5 expression  
 JOURNAL Patent: WO 02053775-A 209 11-JUL-2002;  
 EPIDAURUS BIOTECHNOLOGIE AG (DE)  
 FEATURES  
 source  
 1. .830  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

BASE COUNT 230 a 177 c 237 g 186 t  
 ORIGIN

Query Match 97.1%; Score 33; DB 6; Length 830;  
 Best Local Similarity 100.0%; Pred. No. 0.0034;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATTCTTGTCTACTGGCTGCAGCTGCAGCCCG 33  
 |||||||  
 Db 549 CGATTCTTGTCTACTGGCTGCAGCTGCAGCCCG 581

RESULT 10  
 AX706964  
 LOCUS AX706964 96960 bp DNA PAT 04-APR-2003  
 DEFINITION Sequence 662 from Patent WO03013534.  
 ACCESSION AX706964  
 VERSION AX706964.1 GI:29563289  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 AUTHORS Heinrich,G. and Kerb,R.  
 TITLE Methods for the treatment of cancer with irinotecan based on CYP3A5  
 JOURNAL Patent: WO 03013534-A 662 20-FEB-2003;  
 EPIDAURUS BIOTECHNOLOGIE AG (DE)  
 FEATURES  
 source  
 1. .96960  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 27407 a 19882 c 20450 g 29221 t  
 ORIGIN

Query Match 97.1%; Score 33; DB 6; Length 96960;  
 Best Local Similarity 100.0%; Pred. No. 0.0028;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GATTCCTTGTCTACTGGCTGCAGCTGCAGCCCG 34  
 |||||||  
 Db 93439 GATTCCTTGTCTACTGGCTGCAGCTGCAGCCCG 93471

RESULT 11  
 AX707894  
 LOCUS AX707894 96960 bp DNA PAT 04-APR-2003  
 DEFINITION Sequence 662 from Patent WO03013536.  
 ACCESSION AX707894  
 VERSION AX707894.1 GI:29563969  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 AUTHORS Heinrich,G. and Kerb,R.  
 TITLE Methods for treatment of cancer using irinotecan based on UGT1A1  
 JOURNAL Patent: WO 03013536-A 662 20-FEB-2003;  
 EPIDAURUS BIOTECHNOLOGIE AG (DE)  
 FEATURES  
 source  
 1. .96960  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

BASE COUNT 27407 a 19882 c 20450 g 29221 t  
 ORIGIN

Query Match 97.1%; Score 33; DB 6; Length 96960;  
 Best Local Similarity 100.0%; Pred. No. 0.0028;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GATTCCTTGTCTACTGGCTGCAGCTGCAGCCCG 34  
 |||||||  
 Db 93439 GATTCCTTGTCTACTGGCTGCAGCTGCAGCCCG 93471

RESULT 12  
 AF280107  
 LOCUS AF280107 174832 bp DNA linear PRI 26-MAR-2001  
 DEFINITION Homo sapiens cytochrome P450 polypeptide 43 (CYP3A4) gene, partial  
 cds; cytochrome P450 polypeptide 4 (CYP3A4) and cytochrome P450  
 polypeptide 7 (CYP3A7) genes, complete cds; and cytochrome P450  
 polypeptide 5 (CYP3A5) gene, partial cds.







PA (DNAS-) DNA SCI INC.  
PA (UYDU-) UNIV DUKE.  
XX Guida M, Hall J, Petros WP, Vredenburgh JJ, Colvin OM, Marks JR;  
XX WPI; 2002-691652/74.  
XX New nucleic acid molecule useful for identifying polymorphisms  
PT associated with CYP3A4, CYP3A5 or GSTM1 substrates which influence  
PT breast cancer treatments, comprises at least one base variation from  
PT human CYP3A4 or CYP3A5 sequence -  
XX Claim 9; Page 12; 41pp; English.  
XX The invention relates to a nucleic acid molecule comprising at least one  
CC base variation from human cytochrome P450 (CYP) 3A4 or CYP3A5 sequence.  
CC Nucleic acid molecules of the invention are useful for identifying  
CC polymorphisms associated with CYP3A4, CYP3A5 or glutathione S-transferase  
CC (GST) M1 substrates which influence breast cancer treatments. They are  
CC also useful in diagnostic purposes to identify individuals having a  
CC polymorphic genotype which influence the outcome of breast cancer  
CC treatments and the selection of chemotherapeutic agents used to treat  
CC breast cancer. The polymorphisms detected are used to screen altered  
CC metabolism of CYP3A4, CYP3A5 or GSTM1 substrates, potential drug-drug  
CC interactions, drug adverse effects, likelihood of successful clinical  
CC outcome following treatment with anti-cancer agents such as cisplatin,  
CC cyclophosphamide and/or BCNU. The present sequence is human CYP3A5  
CC polymorphic variant DNA fragment.  
XX Sequence 34 BP; 4 A; 12 C; 9 G; 9 T; 0 other;  
SQ Query Match 100.0%; Score 34; DB 24; Length 34;  
Best Local Similarity 100.0%; Pred. No. 0.00068;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGATTCTTGTCTACTGCTGCAGCTGCAGCCCG 34  
DB 1 CGATTCTTGTCTACTGCTGCAGCTGCAGCCCG 34  
RESULT 2  
ABK99409  
ID ABK99409 standard; DNA; 830 BP.  
XX ABK99409;  
XX 21-OCT-2002 (first entry)  
XX Human CYP3A5 gene polymorphic DNA sequence #1.  
XX Human; CYP3A5; polymorphism; cancer; cardiovascular disease; diabetes;  
KW AIDS; African American; forensic marker; pharmacological; cytostatic;  
KW antidiabetic; anti-HIV; gene therapy; ds.  
XX Homo sapiens.  
XX WO200253775-A2.  
XX 11-JUL-2002.  
XX 21-DEC-2001; 2001WO-EP15290.  
XX 28-DEC-2000; 2000EP-0128627.  
PR 28-DEC-2000; 2000US-258684P.  
PR 29-DEC-2000; 2000US-258952P.  
PR 16-JAN-2001; 2001EP-0100172.  
PR 18-JAN-2001; 2001US-262859P.  
PR 16-AUG-2001; 2001EP-0118884.  
PR 16-AUG-2001; 2001US-312825P.  
XX (EPID-) EPIDAUDOS BIOTECHNOLOGIE AG.  
XX Wojnowski L, Haberl M, Huestert E;  
PI This invention relates to a transgenic non-human animal which may be

XX DR WPI; 2002-583628/62.  
XX Novel CYP3A5 polynucleotide useful for diagnosis and treatment of  
PT cancer, cardiovascular diseases, diabetes and AIDS, and for identifying  
PT polymorphisms -  
XX Claim 1; Figure 4; 138pp; English.  
XX The present invention relates to a new CYP3A5 polynucleotide encoding a  
CC polypeptide, where the polynucleotide is capable of hybridising to a  
CC CYP3A5 gene. The invention is useful in an in vitro method for  
CC identifying a polymorphism. The invention is also useful for useful for  
CC diagnosing a disorder related to the presence of a molecular variant of a  
CC CYP3A5 or susceptibility to such a disorder, where the disorder is  
CC cancer, or diseases including cardiovascular diseases, diabetes and AIDS.  
CC The invention can further be used for the preparation of a diagnostic  
CC composition for diagnosing a disease in a subject having a genome  
CC comprising a variant allele of the CYP3A5 gene, where the subject is an  
CC African American. The molecules of the invention are as forensic markers  
CC and in pharmacological studies. The present nucleic acid sequence  
CC represents a genomic DNA sequence that contains a region in which a  
CC polymorphism of the human CYP3A5 gene was detected.  
XX Sequence 830 BP; 229 A; 176 C; 239 G; 186 T; 0 other;  
SQ Query Match 100.0%; Score 34; DB 24; Length 830;  
Best Local Similarity 100.0%; Pred. No. 0.0012;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGATTCTTGTCTACTGCTGCAGCTGCAGCCCG 34  
DB 549 CGATTCTTGTCTACTGCTGCAGCTGCAGCCCG 582  
RESULT 3  
ABX77171  
ID ABX77171 standard; DNA; 123785 BP.  
XX ABX77171;  
XX 25-APR-2003 (first entry)  
XX DNA sequence of human BAC clone RP11-757A13.  
XX Human; ss; transgenic; drug metabolism; behaviour; gene; mouse;  
KW pharmacokinetic assay; pharmacodynamic assay; toxicology; serum albumin;  
KW alpha-acidic glycoprotein; CYP; multidrug resistance protein; MRP;  
KW uridine diphosphoglucuronosyl transferase; UGT; cytochrome P450.  
XX Homo sapiens.  
XX WO200283897-A1.  
XX 24-OCT-2002.  
XX 18-APR-2002; 2002WO-AU00485.  
XX 18-APR-2001; 2001AU-0004467.  
XX (GENE-) GENE STREAM PTY LTD.  
XX Daly JM;  
XX WPI; 2003-093021/08.  
XX New transgenic non-human animal expressing a foreign polypeptide  
PT associated with drug behavior and/or metabolism, useful for studying  
PT the behavior and/or metabolism of a drug in other animals -  
XX Example 2a; Page 229-295; 408pp; English.  
XX This invention relates to a transgenic non-human animal which may be

used for assessing the behaviour and/or metabolism of a drug in another animal and which expresses a foreign polypeptide associated with drug behaviour and/or metabolism. The invention also comprises a nucleic acid construct for use in producing the above transgenic non-human animal and a method of assessing the metabolism and/or behavior of a drug in an animal of interest, comprising administering a test agent to the transgenic animal and conducting analytical tests to determine drug metabolism and/or behaviour. The transgenic animal is useful in studying drug metabolism and/or behaviour in other animals. The nucleic acid construct is useful in producing the above transgenic animal and the methods are used for producing, breeding and using transgenic animals for pharmacological (e.g. pharmacokinetic or pharmacodynamic assays) and/or toxicological studies. Nucleic acid sequences used within the invention are serum albumin; alpha-acidic glycoprotein; cytochrome P450 (CYP); uridine diphosphoglucuronosyl transferase (UGT); multidrug resistance proteins and (MRP's). The present sequence represents a DNA sequence used to create a transgenic animal within the scope of the invention.

Query Match 100.0%; Score 34; DB 25; Length 123785; Best Local Similarity 100.0%; Pred. No. 0.0027; Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATCTCTTGCTACTGGCTGCGAGTCGAGCCCG 34  
 |||||  
 Db 119180 CGATCTCTTGCTACTGGCTGCGAGTCGAGCCCG 119213

## RESULT 4

ID AAD45763 standard; DNA; 34 BP.

AC AAD45763;

DT 27-DEC-2002 (first entry)

DE Human cytochrome P450 (CYP) 3A5 wild-type DNA fragment.

Human; cytochrome P450; CYP3A4; CYP3A5; glutathione S-transferase; polymorphism; GSTM1; breast cancer; therapy; chemotherapeutic agent; drug-drug interaction; drug adverse effect; anti-cancer agent; enzyme; ds.

OS Homo sapiens.

PN WO200268448-A1.

PD 06-SEP-2002.

PF 26-FEB-2002; 2002WO-US06135.

PR 26-FEB-2001; 2001US-271630P.

PA (DNAS-) DNA SCI INC.

PA (YDU-) UNIV DUKE.

PI Guida M, Hall J, Petros WP, Vredenburg JJ, Colvin OM, Marks JR;

WPI; 2002-691652/74.

PT New nucleic acid molecule useful for identifying polymorphisms associated with CYP3A4, CYP3A5 or GSTM1 substrates which influence breast cancer treatments, comprises at least one base variation from human CYP3A4 or CYP3A5 sequence -

PS Disclosure; Page 12; 41pp; English.

The invention relates to a nucleic acid molecule comprising at least one base variation from human cytochrome P450 (CYP) 3A4 or CYP3A5 sequence. Nucleic acid molecules of the invention are useful for identifying polymorphisms associated with CYP3A4, CYP3A5 or glutathione S-transferase

(GST) M1 substrates which influence breast cancer treatments. They are also useful in diagnostic purposes to identify individuals having a polymorphic genotype which influence the outcome of breast cancer treatments and the selection of chemotherapeutic agents used to treat breast cancer. The polymorphisms detected are used to screen altered metabolism of CYP3A4, CYP3A5 or GSTM1 substrates, potential drug-drug interactions, drug adverse effects, likelihood of successful clinical outcome following treatment with anti-cancer agents such as cisplatin, cyclophosphamide and/or BCNU. The present sequence is human CYP3A5 wild-type DNA fragment.

Query Match 97.1%; Score 33; DB 24; Length 34; Best Local Similarity 100.0%; Pred. No. 0.0017; Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATCTCTTGCTACTGGCTGCGAGTCGAGCCCG 33  
 |||||  
 Db 1 CGATCTCTTGCTACTGGCTGCGAGTCGAGCCCG 33

## RESULT 5

ID ABK99410 standard; DNA; 830 BP.

AC ABK99410;

DT 21-OCT-2002 (first entry)

DE Human CYP3A5 gene polymorphic DNA sequence #2.

Human; CYP3A5; polymorphism; cancer; cardiovascular disease; diabetes; AIDS; African American; forensic marker; pharmacological; cytostatic; antidiabetic; anti-HIV; gene therapy; ds.

OS Homo sapiens.

PN WO200253775-A2.

PD 11-JUL-2002.

PF 21-DEC-2001; 2001WO-EP15290.

PR 28-DEC-2000; 2000EP-0128627.

PR 28-DEC-2000; 2000US-258684P.

PR 29-DEC-2000; 2000US-258952P.

PR 16-JAN-2001; 2001EP-0100172.

PR 18-JAN-2001; 2001US-262859P.

PR 16-AUG-2001; 2001EP-0118884.

PR 16-AUG-2001; 2001US-312825P.

(EPID-) EPIDAURUS BIOTECHNOLOGIE AG.

Wojbowski L, Haberl M, Hustert E;

WPI; 2002-583628/62.

PT Novel CYP3A5 polynucleotide useful for diagnosis and treatment of cancer, cardiovascular diseases, diabetes and AIDS, and for identifying polymorphisms -

PS Claim 1; Figure 4; 138pp; English.

The present invention relates to a new CYP3A5 polynucleotide encoding a polypeptide, where the polynucleotide is capable of hybridising to a CYP3A5 gene. The invention is useful in an in vitro method for identifying a polymorphism. The invention is also useful for useful for diagnosing a disorder related to the presence of a molecular variant of a CYP3A5 or susceptibility to such a disorder, where the disorder is cancer, or diseases including cardiovascular diseases, diabetes and AIDS. The invention can further be used for the preparation of a diagnostic composition for diagnosing a disease in a subject having a genome

comprising a variant allele of the *crp3a5* gene, where the subject is an African American. The molecules of the invention are as forensic markers and in pharmacological studies. The present nucleic acid sequence represents a genomic DNA sequence that contains a region in which a polymorphism of the human *Crp3A5* gene was detected.

Sequence 830 BP; 231 A; 176 C; 236 G; 187 T; 0 other;  
XX SQ

Query Match 97.1%; Score 33; DB 24; Length 830;  
Best Local Similarity 100.0%; Pred.No. 0.0029;  
Matches 33; Conservative 0; Mismatches 0; Indels

Qy 1 CGATTCTTTGCTACTGGTGCAGCTGCAGCCCC 33  
|||||  
Db 549 CGATTCTTTGCTACTGGTGCAGCTGCAGCCCC 58

RESULT 6  
ABK99411  
ID ID ABK99411 standard; DNA; 830 BP.  
XX  
AC ABK99411;

XX	21-OCT-2002	(first entry)
DT		

AA DE Human CYP3A5 gene polymorphic DNA sequence #3.

Human; CYP3A5; polymorphism; cancer; cardiovascular disease; diabetes; AIDS; African American; forensic marker; pharmacological; cytostatic; antidiabetic; anti-HIV; gene therapy; ds.

AA  
OS Homo sapiens.

AA  
PN  
W0200253775-A2.

11-JUL-2002

21-DEC-2001;

PR 28-DEC-2000: 2000EP-0128627.

PR 28-DEC-2000; 2000US-258684P.  
PR 29-DEC-2000; 2000US-258952P.

PR 16-JAN-2001; 2001EP-01001/2.  
PR 18-JAN-2001: 2001US-262859P

PR 16-AUG-2001; 2001EP-0118884.  
PR 16-AUG-2001; 2001US-312825P.

XX  
PA (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.

XX	Wojnowski L.	Haberl M.	Hustert F.
----	--------------	-----------	------------

XX WPI: 2002-583628/62.

XX  
PT  
Novel CYP3A5 polymuc

PT cancer, cardiovascular  
PT polymorphisms -

XX  
PS  
Claim 1: Figure 4: 1

xx The present invention relates to a new CYP3A5 polynucleotide encoding a  
cc polypeptide, where the polynucleotide is capable of hybridizing to a  
cc CYP3A5 gene. The invention is useful in an *in vitro* method for  
cc identifying a polymorphism. The invention is also useful for useful for  
cc diagnosing a disorder related to the presence of a molecular variant of a  
cc CYP3A5 or susceptibility to such a disorder, where the disorder is  
cc cancer, or diseases including cardiovascular diseases, diabetes and AIDS.

The invention can further be used for the preparation of a diagnostic composition for diagnosing a disease in a subject having a genome comprising a variant allele of the CYP3A5 gene, where the subject is an African American. The molecules of the invention are as forensic markers and in pharmacological studies. The present nucleic acid sequence represents a genomic DNA sequence that contains a region in which a polymorphism of the human CYP3A5 gene was detected.

```

XX
SQ  Sequence 830 BP; 231 A; 176 C; 236 G; 187 T; 0 other;

Query Match      97.1%; Score 33; DB 24; Length 830;
Best Local Similarity 100.0%; Pred.No. 0.0029;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 CGATTCTTTTGCTACTGGCTGCAGTGCAGCCCC 33  
.  
|||||  
549 CGATTCTTTTGCTACTGGCTGCAGTGCAGCCCC 581

RESULT 7  
ABK99499  
ID ABK99499 standard; DNA; 830 BP.  
XX  
AC ABK99499.

21-OCT-2002 (first entry)

Human CYP3A5 gene polymorphic DNA sequence #30.

Human; CYP3A5; polymorphism; cancer; cardiovascular disease; diabetes;  
 AIDS; African American; forensic marker; pharmacological; cytostatic;  
 antidiabetic; anti-HIV; gene therapy; ds.

XX Homo sapiens. OS

XX PN WO200253775-A2.

XX  
PD  
11-JUL-2002XX  
PF  
21-DEC-2001:XX  
PR 28-DEC-2000: 2000EP-0128627.

PR 28-DEC-2000; 2000US-258684P.  
PR 29-DEC-2000; 2000US-258952P

PR 16-JAN-2001; 2001EP-0100172.  
PR 18-JAN-2001; 2001US-262859P.

PR 16-AUG-2001; 2001EP-0118884.  
PR 16-AUG-2001; 2001US-312825P

XX  
PA (EPID-) EPIDAUROS BIOTECHNOIXX  
PT Wojnowski T., Haberl M., Husterst F:XX  
DR WPI: 2002-583628/62.

AA	
PT	Novel CYP3A5 polynucleotide useful for diagnosis and treatment of
PT	cancer, cardiovascular diseases, diabetes and AIDS, and for identifying
PT	polymorphisms -

XX  
PS  
Claim 1: Figure 4: 138pp: English:

xx The present invention relates to a new CYP3A5 polynucleotide encoding a  
cc polypeptide, where the polynucleotide is capable of hybridising to a  
cc CYP3A5 gene. The invention is useful in an in vitro method for  
cc identifying a polymorphism. The invention is also useful for useful for  
cc diagnosing a disorder related to the presence of a molecular variant of a  
cc CYP3A5 or susceptibility to such a disorder, where the disorder is  
cc cancer, or diseases including cardiovascular diseases, diabetes and AIDS.  
cc The invention can further be used for the preparation of a diagnostic  
cc composition for diagnosing a disease in a subject having a genome  
cc comprising a variant allele of the CYP3A5 gene, where the subject is an  
cc African American. The molecules of the invention are as forensic markers  
cc and in pharmacological studies. The present nucleic acid sequence  
cc represents a genomic DNA sequence that contains a region in which a  
cc polymorphism of the human CYP3A5 gene was detected.

Sequence 830 BP: 230 A: 177 C: 237 G: 186 T: 0 other: 0

Query Match 97.1%; Score 33; DB 24; Length 830;  
Best Local Similarity 100.0%; Pred. No. 0.0029;



Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATCTTTGCTACTGGCTGCAGCTGCAGCCCC 33  
|||||  
Db 549 CGATCTTTGCTACTGGCTGCAGCTGCAGCCCC 581

## RESULT 8

ABK99500  
ID ABK99500 standard; DNA; 830 BP.  
XX  
AC ABK99500;  
XX  
DT 21-OCT-2002 (first entry)  
XX  
DE Human CYP3A5 gene polymorphic DNA sequence #31.  
XX  
KW Human; CYP3A5; polymorphism; cancer; cardiovascular disease; diabetes;  
KW AIDS; African American; forensic marker; pharmacological; cytostatic;  
KW antidiabetic; anti-HIV; gene therapy; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200253775-A2.  
XX  
PD 11-JUL-2002.  
XX  
PF 21-DEC-2001; 2001WO-EP15290.  
XX  
PR 28-DEC-2000; 2000EP-0128627.  
PR 28-DEC-2000; 2000US-258684P.  
PR 29-DEC-2000; 2000US-258952P.  
PR 16-JAN-2001; 2001EP-0100172.  
PR 18-JAN-2001; 2001US-262859P.  
PR 16-AUG-2001; 2001EP-0118884.  
PR 16-AUG-2001; 2001US-312825P.  
XX  
PA (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.  
XX  
PI Wojnowski L, Haberl M, Hustert E;  
XX  
DR WPI; 2002-583628/62.  
XX

## Claim 1; Figure 4; 138pp; English.

The present invention relates to a new CYP3A5 polynucleotide encoding a polypeptide, where the polynucleotide is capable of hybridizing to a CYP3A5 gene. The invention is useful in an in vitro method for identifying a polymorphism. The invention is also useful for useful for diagnosing a disorder related to the presence of a molecular variant of a CYP3A5 or susceptibility to such a disorder, where the disorder is cancer, or diseases including cardiovascular diseases, diabetes and AIDS. The invention can further be used for the preparation of a diagnostic comprising a variant allele of the CYP3A5 gene, where the subject is an African American. The molecules of the invention are as forensic markers and in pharmacological studies. The present nucleic acid sequence represents a genomic DNA sequence that contains a region in which a polymorphism of the human CYP3A5 gene was detected.

Query Match 97.1%; Score 33; DB 24; Length 830;

Best Local Similarity 100.0%; Pred. No. 0.0029;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATCTTTGCTACTGGCTGCAGCTGCAGCCCC 33  
|||||  
Db 549 CGATCTTTGCTACTGGCTGCAGCTGCAGCCCC 581

## RESULT 9

ABK99501  
ID ABK99501 standard; DNA; 830 BP.  
XX  
AC ABK99501;  
XX  
DT 21-OCT-2002 (first entry)  
XX  
DE Human CYP3A5 gene polymorphic DNA sequence #32.  
XX  
KW Human; CYP3A5; polymorphism; cancer; cardiovascular disease; diabetes;  
KW AIDS; African American; forensic marker; pharmacological; cytostatic;  
KW antidiabetic; anti-HIV; gene therapy; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200253775-A2.  
XX  
PD 11-JUL-2002.  
XX  
PF 21-DEC-2001; 2001WO-EP15290.  
XX  
PR 28-DEC-2000; 2000EP-0128627.  
PR 28-DEC-2000; 2000US-258684P.  
PR 29-DEC-2000; 2000US-258952P.  
PR 16-JAN-2001; 2001EP-0100172.  
PR 18-JAN-2001; 2001US-262859P.  
PR 16-AUG-2001; 2001EP-0118884.  
PR 16-AUG-2001; 2001US-312825P.  
XX  
PA (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.  
XX  
PI Wojnowski L, Haberl M, Hustert E;  
XX  
DR WPI; 2002-583628/62.  
XX

## Claim 1; Figure 4; 138pp; English.

The present invention relates to a new CYP3A5 polynucleotide encoding a polypeptide, where the polynucleotide is capable of hybridizing to a CYP3A5 gene. The invention is useful in an in vitro method for identifying a polymorphism. The invention is also useful for useful for diagnosing a disorder related to the presence of a molecular variant of a CYP3A5 or susceptibility to such a disorder, where the disorder is cancer, or diseases including cardiovascular diseases, diabetes and AIDS. The invention can further be used for the preparation of a diagnostic comprising a variant allele of the CYP3A5 gene, where the subject is an African American. The molecules of the invention are as forensic markers and in pharmacological studies. The present nucleic acid sequence represents a genomic DNA sequence that contains a region in which a polymorphism of the human CYP3A5 gene was detected.

Query Match 97.1%; Score 33; DB 24; Length 830;

Best Local Similarity 100.0%; Pred. No. 0.0029;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATCTTTGCTACTGGCTGCAGCTGCAGCCCC 33  
|||||  
Db 549 CGATCTTTGCTACTGGCTGCAGCTGCAGCCCC 581

## RESULT 10

AAA51756  
ID AAA51756 standard; DNA; 1346 BP.

XX AC AA51756;  
 XX 31-OCT-2000 (first entry)  
 XX Cytochrome P450 CYP3A5 gene 5' flanking region (-1343 to +3).  
 DE CYP3A5; Cytochrome P450; transcription regulatory region; polymorphism;  
 KW Activator protein-3 motif; AP-3; basic transcription element;  
 KW drug metabolism; phenotype; ss.  
 XX Homo sapiens.  
 OS  
 XX Key Location/Qualifiers  
 FT replace(869,G)  
 FT /tag= a  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT /note= "Corresponds to nucleotide position -475. Presence  
 FT of a G in variants destroys an AP-3 site"  
 FT variation  
 FT replace(1197,G)  
 FT /tag= b  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT /note= "Corresponds to nucleotide position -147. Presence  
 FT of a G in variants destroys a basic transcription  
 FT element"  
 XX WO2000039332-A1.  
 PN 06-JUL-2000.  
 PD 22-DEC-1999; 99WO-GB04380.  
 PF 23-DEC-1998; 98GB-0028619.  
 PR (JANC ) JANSSEN PHARM NV.  
 PA Paulussen ADC, Armstrong M;  
 PI WPI; 2000-452418/39.  
 XX Identifying subjects with a high drug metabolizing phenotype associated  
 PT with cytochrome CYP3A5 expression for establishing whether a drug will  
 PT be metabolized by the subject  
 XX Claim 5; Fig 7; 68pp; English.  
 PS Cytochrome P450 subfamily CYP3A5 transcription regulatory regions can be  
 CC screened for the presence/absence of a polymorphic variant, preferably  
 CC at positions -475 or -147 of the DNA of the 5' flanking region adjacent  
 CC to the CYP3A5 coding sequence. The variants are present in an activator  
 CC protein-3 (AP-3) motif and/or a basic transcription element (BRE). The  
 CC polymorphisms cause increased CYP3A5 gene expression and this has been  
 CC linked to drug metabolic activity. Screening for the presence of variants  
 CC can be used to identify subjects with a high or low drug metabolizing  
 CC phenotype associated with cytochrome CYP3A5 expression. Primers are used  
 CC which in addition to hybridizing to the site of interest, are capable of  
 CC introducing a restriction site which is absent in either the wild type  
 CC sequence or polymorphic variants. Restriction enzyme cleavage analysis  
 CC can then be used to indicate the presence or absence of the variant. The  
 CC methods are used to establish, before treatment with a drug, whether the  
 CC drug will be effectively metabolized by the patient, to identify  
 CC compounds and transcription factors that can bind to a DNA sequence  
 CC encoding CYP3A5, diagnosing susceptibility to a disease which is caused  
 CC by toxins or procarcinogens metabolized by CYP3A5 and for identifying  
 CC mutagenic effects of a compound.  
 XX Sequence 1346 BP; 409 A; 275 C; 344 G; 318 T; 0 other;  
 SQ Query Match 97.1%; Score 33; DB 21; Length 1346;  
 Best Local Similarity 100.0%; Pred. No. 0.0031;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CGATTCCTTGTCTACTGGCTGCAGCTGCAGCCCC 33

DB 1164 CGATTCCTTGTCTACTGGCTGCAGCTGCAGCCCC 1196  
 RESULT 11  
 ABK99412  
 ID ABK99412 standard; DNA; 830 BP.  
 XX AC ABK99412;  
 XX 21-OCT-2002 (first entry)  
 XX Human CYP3A5 gene polymorphic DNA sequence #4.  
 DE Human; CYP3A5; polymorphism; cancer; cardiovascular disease; diabetes;  
 XX AIDS; African American; forensic marker; pharmacological; cytostatic;  
 KW antidiabetic; anti-HIV; gene therapy; ds.  
 XX Homo sapiens.  
 OS WO200253775-A2.  
 XX PN 11-JUL-2002.  
 PD 21-DEC-2001; 2001WO-BP15290.  
 PF 28-DEC-2000; 2000EP-0128627.  
 PR 28-DEC-2000; 2000US-258684P.  
 PR 29-DEC-2000; 2000US-258952P.  
 PR 16-JAN-2001; 2001EP-0100172.  
 PR 18-JAN-2001; 2001US-262859P.  
 PR 16-AUG-2001; 2001EP-0118884.  
 PR 16-AUG-2001; 2001US-312825P.  
 XX (EPID-) EPIDAURUS BIOTECHNOLOGIE AG.  
 XX PI Wojnowski L, Haberl M, Hustert E;  
 XX WPI; 2002-583628/62.  
 DR Novel CYP3A5 polynucleotide useful for diagnosis and treatment of  
 PT cancer, cardiovascular diseases, diabetes and AIDS, and for identifying  
 PT polymorphisms.  
 XX Claim 1; Figure 4; 138pp; English.  
 PS The present invention relates to a new CYP3A5 polynucleotide encoding a  
 CC polypeptide, where the polynucleotide is capable of hybridizing to a  
 CC CYP3A5 gene. The invention is useful in an in vitro method for  
 CC identifying a polymorphism. The invention is also useful for useful for  
 CC diagnosing a disorder related to the presence of a molecular variant of a  
 CC CYP3A5 or susceptibility to such a disorder, where the disorder is  
 CC cancer, or diseases including cardiovascular diseases, diabetes and AIDS.  
 CC The invention can further be used for the preparation of a diagnostic  
 CC composition for diagnosing a disease in a subject having a genome  
 CC comprising a variant allele of the CYP3A5 gene, where the subject is an  
 CC African American. The molecules of the invention are as forensic markers  
 CC and in pharmacological studies. The present nucleic acid sequence  
 CC represents a genomic DNA sequence that contains a region in which a  
 CC polymorphism of the human CYP3A5 gene was detected.  
 XX Sequence 830 BP; 231 A; 176 C; 236 G; 187 T; 0 other;  
 SQ Query Match 92.4%; Score 31.4; DB 24; Length 830;  
 Best Local Similarity 97.0%; Pred. No. 0.012;  
 Matches 32; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CGATTCCTTGTCTACTGGCTGCAGCTGCAGCCCC 33  
 DB 549 CAATTCCTTGTCTACTGGCTGCAGCTGCAGCCCC 581  
 RESULT 12

ABK99524  
ID ABK99524 standard; DNA; 830 BP.

XX AC  
XX ABK99524;

DT 21-OCT-2002 (first entry)

XX Human CYP3A5 gene polymorphic DNA sequence #55.

XX Human; CYP3A5; polymorphism; cancer; cardiovascular disease; diabetes;  
KW AIDS; African American; forensic marker; pharmacological; cytostatic;  
KW antidiabetic; anti-HIV; gene therapy; ds.

XX OS Homo sapiens.

XX PN WO200253775-A2.

XX 11-JUL-2002.

PF 21-DEC-2001; 2001WO-EP15290.

XX 28-DEC-2000; 2000EP-0128627.

PR 28-DEC-2000; 2000US-258684P.

PR 29-DEC-2000; 2000US-258952P.

PR 16-JAN-2001; 2001EP-0100172.

PR 18-JAN-2001; 2001US-262859P.

PR 16-AUG-2001; 2001EP-0118884.

PR 16-AUG-2001; 2001US-312825P.

XX (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.

XX WOjnowski L, Haberl M, Hustert E;

XX WPI; 2002-583628/62.

XX Novel CYP3A5 polynucleotide useful for diagnosis and treatment of  
PT cancer, cardiovascular diseases, diabetes and AIDS, and for identifying  
PT polymorphisms -

XX Claim 1; Figure 4; 138pp; English.

XX The present invention relates to a new CYP3A5 polynucleotide encoding a  
CC polypeptide, where the polynucleotide is capable of hybridising to a  
CC CYP3A5 gene. The invention is useful in an in vitro method for  
CC identifying a polymorphism. The invention is also useful for useful for  
CC diagnosing a disorder related to the presence of a molecular variant of a  
CC CYP3A5 or susceptibility to such a disorder, where the disorder is  
CC cancer, or diseases including cardiovascular diseases, diabetes and AIDS.  
CC The invention can further be used for the preparation of a diagnostic  
CC composition for diagnosing a disease in a subject having a genome  
CC comprising a variant allele of the CYP3A5 gene, where the subject is an  
CC African American. The molecules of the invention are as forensic markers  
CC and in pharmacological studies. The present nucleic acid sequence  
CC represents a genomic DNA sequence that contains a region in which a  
CC polymorphism of the human CYP3A5 gene was detected.

XX Sequence 830 BP; 230 A; 177 C; 237 G; 186 T; 0 other;

Query Match 92.4%; Score 31.4; DB 24; Length 830;

Best Local Similarity 97.0%; Pred. No. 0.012;

Matches 32; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGATTCTTTGCTACTGCTGCAGTCGACGCC 33

DB 549 CGATTCTTTGCTACTGCTGCAGTCGACGCC 581

RESULT 13

ABK99421

ID ABK99421 standard; DNA; 621 BP.

XX AC  
XX ABK99421;

DT 21-OCT-2002 (first entry)

XX Human CYP3A5 gene polymorphic DNA sequence #13.

XX Human; CYP3A5; polymorphism; cancer; cardiovascular disease; diabetes;  
KW AIDS; African American; forensic marker; pharmacological; cytostatic;  
KW antidiabetic; anti-HIV; gene therapy; ds.

XX OS Homo sapiens.

XX PN WO200253775-A2.

XX 11-JUL-2002.

XX 21-DEC-2001; 2001WO-EP15290.

XX 28-DEC-2000; 2000EP-0128627.

PR 28-DEC-2000; 2000US-258684P.

PR 29-DEC-2000; 2000US-258952P.

PR 16-JAN-2001; 2001EP-0100172.

PR 18-JAN-2001; 2001US-262859P.

PR 16-AUG-2001; 2001EP-0118884.

PR 16-AUG-2001; 2001US-312825P.

XX (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.

XX WOjnowski L, Haberl M, Hustert E;

XX WPI; 2002-583628/62.

XX Novel CYP3A5 polynucleotide useful for diagnosis and treatment of  
PT cancer, cardiovascular diseases, diabetes and AIDS, and for identifying  
PT polymorphisms -

XX Claim 1; Figure 4; 138pp; English.

XX The present invention relates to a new CYP3A5 polynucleotide encoding a  
CC polypeptide, where the polynucleotide is capable of hybridising to a  
CC CYP3A5 gene. The invention is useful in an in vitro method for  
CC identifying a polymorphism. The invention is also useful for useful for  
CC diagnosing a disorder related to the presence of a molecular variant of a  
CC CYP3A5 or susceptibility to such a disorder, where the disorder is  
CC cancer, or diseases including cardiovascular diseases, diabetes and AIDS.  
CC The invention can further be used for the preparation of a diagnostic  
CC composition for diagnosing a disease in a subject having a genome  
CC comprising a variant allele of the CYP3A5 gene, where the subject is an  
CC African American. The molecules of the invention are as forensic markers  
CC and in pharmacological studies. The present nucleic acid sequence  
CC represents a genomic DNA sequence that contains a region in which a  
CC polymorphism of the human CYP3A5 gene was detected.

XX Sequence 621 BP; 172 A; 134 C; 148 G; 167 T; 0 other;

Query Match 81.2%; Score 27.6; DB 24; Length 621;

Best Local Similarity 88.2%; Pred. No. 0.34;

Matches 30; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGATTCTTTGCTACTGCTGCAGTCGACGCC 34

DB 65 CGATTCTTTGCTACTGCTGCAGTCGACGCC 98

RESULT 14

ABK99419

ID ABK99419 standard; DNA; 624 BP.

XX AC  
XX ABK99419;

DT 21-OCT-2002 (first entry)

XX Human CYP3A5 gene polymorphic DNA sequence #11.

XX Human; CYP3A5; polymorphism; cancer; cardiovascular disease; diabetes;

KW AIDS; African American; forensic marker; pharmacological; cytostatic;  
 KW antidiabetic; anti-HIV; gene therapy; ds.  
 XX  
 XX Homo sapiens.

OS

PN WO200253775-A2.

XX 11-JUL-2002.

PD

PF 21-DEC-2001; 2001WO-EP15290.

XX 28-DEC-2000; 2000EP-0128627.

PR 28-DEC-2000; 2000US-258684P.

PR 29-DEC-2000; 2000US-258952P.

PR 16-JAN-2001; 2001EP-0100172.

PR 18-JAN-2001; 2001US-262859P.

PR 16-AUG-2001; 2001EP-0118884.

PR 16-AUG-2001; 2001US-312825P.

XX

PA (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.

XX

PI Wojnowski L, Haberl M, Huestert E;

XX

DR WPI; 2002-583628/62.

XX

PS Claim 1; Figure 4; 138pp; English.

XX

CC The present invention relates to a new CYP3A5 polynucleotide encoding a

CC polypeptide, where the polynucleotide is capable of hybridizing to a

CC CYP3A5 gene. The invention is useful in an in vitro method for

CC identifying a polymorphism. The invention is also useful for useful for

CC diagnosing a disorder related to the presence of a molecular variant of a

CC CYP3A5 or susceptibility to such a disorder, where the disorder is

CC cancer, or diseases including cardiovascular diseases, diabetes and AIDS.

CC The invention can further be used for the preparation of a diagnostic

CC composition for diagnosing a disease in a subject having a genome

CC comprising a variant allele of the CYP3A5 gene, where the subject is an

CC African American. The molecules of the invention are as forensic markers

CC and in pharmacological studies. The present nucleic acid sequence

CC represents a genomic DNA sequence that contains a region in which a

CC polymorphism of the human CYP3A5 gene was detected.

XX

SQ Sequence 624 BP; 174 A; 134 C; 149 G; 167 T; 0 other;

Query Match 81.2%; Score 27.6; DB 24; Length 624;  
 Best Local Similarity 88.2%; Pred. No. 0.34;  
 Matches 30; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGATTCTTTGCTACTGCTGCAGCTGCAGCCCG 34

Db 65 CGATTCTTTGCTATTGGCTGCAGCTATAGCCCTG 98

RESULT 15

ABK99420

ID ABK99420 standard; DNA; 624 BP.

XX

AC ABK99420;

XX

DT 21-OCT-2002. (first entry)

XX

DE Human CYP3A5 gene polymorphic DNA sequence #12.

XX

KW Human; CYP3A5; polymorphism; cancer; cardiovascular disease; diabetes;

KW AIDS; African American; forensic marker; pharmacological; cytostatic;

KW antidiabetic; anti-HIV; gene therapy; ds.

XX

OS Homo sapiens.

XX

PN WO200253775-A2.

XX 11-JUL-2002.

XX

PF 21-DEC-2001; 2001WO-EP15290.

XX

PR 28-DEC-2000; 2000EP-0128627.

PR 28-DEC-2000; 2000US-258684P.

PR 29-DEC-2000; 2000US-258952P.

PR 16-JAN-2001; 2001EP-0100172.

PR 18-JAN-2001; 2001US-262859P.

PR 16-AUG-2001; 2001EP-0118884.

PR 16-AUG-2001; 2001US-312825P.

XX

PA (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.

XX

PI Wojnowski L, Haberl M, Huestert E;

XX

DR WPI; 2002-583628/62.

XX

PS Claim 1; Figure 4; 138pp; English.

XX

CC The present invention relates to a new CYP3A5 polynucleotide encoding a

CC polypeptide, where the polynucleotide is capable of hybridizing to a

CC CYP3A5 gene. The invention is useful in an in vitro method for

CC identifying a polymorphism. The invention is also useful for useful for

CC diagnosing a disorder related to the presence of a molecular variant of a

CC CYP3A5 or susceptibility to such a disorder, where the disorder is

CC cancer, or diseases including cardiovascular diseases, diabetes and AIDS.

CC The invention can further be used for the preparation of a diagnostic

CC composition for diagnosing a disease in a subject having a genome

CC comprising a variant allele of the CYP3A5 gene, where the subject is an

CC African American. The molecules of the invention are as forensic markers

CC and in pharmacological studies. The present nucleic acid sequence

CC represents a genomic DNA sequence that contains a region in which a

CC polymorphism of the human CYP3A5 gene was detected.

XX

SQ Sequence 624 BP; 173 A; 133 C; 150 G; 168 T; 0 other;

Query Match 81.2%; Score 27.6; DB 24; Length 624;  
 Best Local Similarity 88.2%; Pred. No. 0.34;  
 Matches 30; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGATTCTTTGCTACTGCTGCAGCTGCAGCCCG 34

Db 65 CGATTCTTTGCTATTGGCTGCAGCTATAGCCCTG 98

Search completed: August 27, 2003, 18:15:45

Job time : 265.808 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 27, 2003, 18:02:39 ; Search time 37.0385 Seconds  
(without alignments)  
214.504 Million cell updates/sec

Title: US-10-085-612-1

Perfect score: 18

Sequence: 1 gacaaaggcagcagag 18

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued\_Patents\_NA:\*
- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*
  - 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*
  - 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*
  - 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*
  - 5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq:\*
  - 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
c 1	16.4	91.1	29	4	US-09-144-367-13
c 2	16.4	91.1	1345	3	US-09-372-339-2
c 3	16	88.9	1345	4	US-09-144-367-3
c 4	15.4	85.6	19	4	US-09-144-367-44
c 5	15.4	85.6	45546	4	US-09-146-053-6
c 6	14.8	82.2	32	4	US-09-144-367-12
c 7	14.8	82.2	807	1	US-08-154-019-28
c 8	14.8	82.2	807	1	US-08-461-333-28
c 9	14.8	82.2	807	3	US-08-464-167-28
c 10	14.8	82.2	807	3	US-09-158-313-28
c 11	14.8	82.2	807	3	US-08-476-798-28
c 12	14.8	82.2	824	1	US-08-154-019-29
c 13	14.8	82.2	824	1	US-08-461-333-29
c 14	14.8	82.2	824	3	US-08-464-167-29
c 15	14.8	82.2	824	3	US-09-158-313-29
c 16	14.8	82.2	824	3	US-08-476-798-29
c 17	14.8	82.2	1202	4	US-09-904-615-63
c 18	14.8	82.2	1345	3	US-09-372-339-1
c 19	14.8	82.2	1452	4	US-09-904-615-34
c 20	14.8	82.2	1950	4	US-09-205-258-131
c 21	14.8	82.2	4523	4	US-09-473-716-1
c 22	14.8	82.2	6763	2	US-08-756-506-23
c 23	14.8	82.2	10807	1	US-08-206-176-7
c 24	14.8	82.2	10807	2	US-08-756-506-5
c 25	14.8	82.2	16063	4	US-09-801-052-3
c 26	14.8	82.2	29629	4	US-09-729-995-3
c 27	14.8	82.2	48974	3	US-08-920-422-17

28	14.8	82.2	152331	3	US-09-128-155-16	Sequence 16, Appl
29	14.8	80.0	176373	3	US-09-128-155-17	Sequence 17, Appl
30	14.4	82.0	19	3	US-09-177-359-37	Sequence 37, Appl
c 31	14.4	80.0	420	4	US-09-702-705-1114	Sequence 1114, Ap
c 32	14.4	80.0	420	4	US-09-736-457-1114	Sequence 1114, Ap
c 33	14.4	80.0	732	4	US-09-149-476-66	Sequence 66, Appl
c 34	14.4	80.0	899	4	US-09-171-209-50	Sequence 50, Appl
c 35	14.4	80.0	2043	4	US-09-149-476-231	Sequence 231, App
c 36	14.4	80.0	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c 37	14.4	80.0	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c 38	14	77.8	4267	3	US-08-949-155-51	Sequence 51, Appl
c 39	14	77.8	4267	4	US-09-819-964-51	Sequence 51, Appl
c 40	14	77.8	35100	1	US-08-306-691B-19	Sequence 19, Appl
c 41	14	77.8	35100	5	PCT-US93-06251-19	Sequence 19, Appl
c 42	13.8	76.7	19	4	US-09-144-367-43	Sequence 43, Appl
c 43	13.8	76.7	114	3	US-09-172-841-50	Sequence 50, Appl
c 44	13.8	76.7	114	4	US-08-951-621-50	Sequence 50, Appl
c 45	13.8	76.7	990	2	US-08-688-342-2	Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-09-144-367-13/c  
; Sequence 13, Application US/09144367  
; Patent No. 6432639  
; GENERAL INFORMATION:  
; APPLICANT: Lichter, Jay  
; APPLICANT: Guido, Marco  
; TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4  
; FILE REFERENCE: SEQ-12P  
; CURRENT APPLICATION NUMBER: US/09/144,367  
; CURRENT FILING DATE: 1998-08-31  
; PRIOR APPLICATION NUMBER: 60/058,612  
; PRIOR FILING DATE: 1997-09-10  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 13  
; LENGTH: 29  
; TYPE: DNA  
; ORGANISM: H. sapiens  
US-09-144-367-13

Query Match 91.1%; Score 16.4; DB 4; Length 29;  
Best Local Similarity 94.4%; Pred. No. 24; Indels 0; Gaps 0;  
Matches 17; Conservative 0; Mismatches 1;

QY 1 GACAAGGCGCAGCAGAG 18  
|||||  
Db 24 GACAAGGCGCAGCAGAG 7

RESULT 2  
US-09-372-339-2  
; Sequence 2, Application US/09372339  
; Patent No. 6174684  
; GENERAL INFORMATION:  
; APPLICANT: Rebbeck, Timothy  
; APPLICANT: Felix, Carolyn  
; TITLE OF INVENTION: CYP3A4 NFSE Variant and Methods of Use Therefor  
; FILE REFERENCE: PENN-0695  
; CURRENT APPLICATION NUMBER: US/09/372,339  
; CURRENT FILING DATE: 1999-08-11  
; EARLIER APPLICATION NUMBER: 60/096,586  
; EARLIER FILING DATE: 1998-08-14  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 1345  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-372-339-2



;; TITLE OF INVENTION: Production of Recombinant Polypeptides  
;; TITLE OF INVENTION: by Bovine Species and Transgenic Methods  
;; NUMBER OF SEQUENCES: 38

## CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Townsend and Townsend and Crew  
;; STREET: One Market Plaza, Steuart Tower, Suite 2000  
;; CITY: San Francisco  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 94105

## COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25

## CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/154,019  
;; FILING DATE: 16-NOV-1993  
;; CLASSIFICATION: 435

## PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 08/461,333  
;; FILING DATE: 05-JUN-1995  
;; APPLICATION NUMBER: US 08/077,788  
;; FILING DATE: 15-JUN-1993

## PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/895,956  
;; FILING DATE: 15-JUN-1992  
;; APPLICATION NUMBER: US 07/619,131  
;; FILING DATE: 27-NOV-1990

## PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/444,745  
;; FILING DATE: 01-DEC-1989  
;; ATTORNEY/AGENT INFORMATION:

;; NAME: Liebescheutz, Joe O.  
;; REGISTRATION NUMBER: 37,505

;; REFERENCE/DOCKET NUMBER: 16994-003123

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 415-543-9600

;; TELEFAX: 415-543-5043

;; INFORMATION FOR SEQ ID NO: 28:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 807 base pairs

;; TYPE: nucleic acid

;; STRANDEDNESS: single

;; TOPOLOGY: linear

;; MOLECULE TYPE: DNA (genomic)

;; US-08-154-019-28

Query Match 82.2%; Score 14.8; DB 1; Length 807;

Best Local Similarity 88.9%; Pred. No. 1.9e+02;

Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GACAAGGCGAGCAGACAG 18

Db 411 GACAAGGCGAGCAGACAG 394

RESULT 8

US-08-461-333-28/c

Sequence 28, Application US/08461333

Patent No. 5741957

GENERAL INFORMATION:

APPLICANT: Deboer, Herman A.

APPLICANT: Strijker, Rein

APPLICANT: Heyneker, Herbert L.

APPLICANT: Platenburg, Gerald

APPLICANT: Lee, Sang He

APPLICANT: Pieper, Frank

APPLICANT: Krimpenfort, Paul J.A.

TITLE OF INVENTION: Production of Recombinant Polypeptides

TITLE OF INVENTION: by Bovine Species and Transgenic Methods

NUMBER OF SEQUENCES: 38

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

## CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Townsend and Townsend and Crew  
;; STREET: One Market Plaza, Steuart Tower, Suite 2000  
;; CITY: San Francisco  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 94105

## COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25

## CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/461,333  
;; FILING DATE: 05-JUN-1995  
;; CLASSIFICATION: 800

## PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 08/077,788  
;; FILING DATE: 15-JUN-1993

## PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/895,956  
;; FILING DATE: 15-JUN-1992  
;; APPLICATION NUMBER: US 07/619,131  
;; FILING DATE: 27-NOV-1990

## PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/444,745  
;; FILING DATE: 01-DEC-1989  
;; ATTORNEY/AGENT INFORMATION:

;; NAME: Liebescheutz, Joe O.  
;; REGISTRATION NUMBER: 37,505

;; REFERENCE/DOCKET NUMBER: 16994-003123

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 415-543-9600

;; TELEFAX: 415-543-5043

;; INFORMATION FOR SEQ ID NO: 28:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 807 base pairs

;; TYPE: nucleic acid

;; STRANDEDNESS: single

;; TOPOLOGY: linear

;; MOLECULE TYPE: DNA (genomic)

;; US-08-461-333-28

Query Match 82.2%; Score 14.8; DB 1; Length 807;

Best Local Similarity 88.9%; Pred. No. 1.9e+02;

Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GACAAGGCGAGCAGACAG 18

Db 411 GACAAGGCGAGCAGACAG 394

RESULT 9

US-08-461-167-28/c

Sequence 28, Application US/0846167

Patent No. 6013857

GENERAL INFORMATION:

APPLICANT: Deboer, Herman A.

APPLICANT: Strijker, Rein

APPLICANT: Heyneker, Herbert L.

APPLICANT: Platenburg, Gerald

APPLICANT: Lee, Sang He

APPLICANT: Pieper, Frank

APPLICANT: Krimpenfort, Paul J.A.

TITLE OF INVENTION: Production of Recombinant Polypeptides

TITLE OF INVENTION: by Bovine Species and Transgenic Methods

NUMBER OF SEQUENCES: 38

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

;;

```

; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 05-JUN-1995
; APPLICATION NUMBER: US/08/464,167
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/077,788
; FILING DATE: 15-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/895,956
; FILING DATE: 15-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/619,131
; FILING DATE: 27-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/444,745
; FILING DATE: 01-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Liebescheutz, Joe O.
; REGISTRATION NUMBER: 37,505
; REFERENCE/DOCKET NUMBER: 16994-003124
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 807 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-464-167-28

Query Match      82.2%; Score 14.8; DB 3; Length 807;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GACAGGCGGAGGACAGAG 18
Db 411 GACAGGCGGAGGTCAGAG 394

RESULT 10
US-09-158-313-28/c
; Sequence 28, Application US/09158313
; Patent No 6066725
; GENERAL INFORMATION:
; APPLICANT: Deboer, Herman A.
; APPLICANT: Strijker, Rein
; APPLICANT: Heyneker, Herbert L.
; APPLICANT: Platenburg, Gerald
; APPLICANT: Lee, Sang He
; APPLICANT: Pieper, Frank
; APPLICANT: Krimpenfort, Paul J.A.
; TITLE OF INVENTION: Production of Recombinant Polypeptides
; TITLE OF INVENTION: by Bovine Species and Transgenic Methods
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/158,313
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,798
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/895,956
; FILING DATE: 15-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/619,131
; FILING DATE: 27-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/444,745
; FILING DATE: 01-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Liebescheutz, Joe O.
; REGISTRATION NUMBER: 37,505
; REFERENCE/DOCKET NUMBER: 16994-003125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 807 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-158-313-28

Query Match      82.2%; Score 14.8; DB 3; Length 807;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GACAGGCGGAGGACAGAG 18
Db 411 GACAGGCGGAGGTCAGAG 394

RESULT 11
US-08-476-798-28/c
; Sequence 28, Application US/08476798
; Patent No. 6140552
; GENERAL INFORMATION:
; APPLICANT: Deboer, Herman A.
; APPLICANT: Strijker, Rein
; APPLICANT: Heyneker, Herbert L.
; APPLICANT: Platenburg, Gerald
; APPLICANT: Lee, Sang He
; APPLICANT: Pieper, Frank
; APPLICANT: Krimpenfort, Paul J.A.
; TITLE OF INVENTION: Production of Recombinant Polypeptides
; TITLE OF INVENTION: by Bovine Species and Transgenic Methods
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,798
; FILING DATE: 07-JUN-1995

```



```

; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/077,788
; FILING DATE: 15-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/895,956
; FILING DATE: 15-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/619,131
; FILING DATE: 27-NOV-1990
; APPLICATION NUMBER: US 07/444,745
; FILING DATE: 01-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/444,745
; FILING DATE: 01-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Liebescheutz, Joe O.
; REGISTRATION NUMBER: 37,505
; REFERENCE/DOCKET NUMBER: 16994-003125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 807 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-476-798-28

```

```

Query Match      82.2%; Score 14.8; DB 3; Length 807;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 GACAGGGCAGCAGCAGAG 18
    ||||| |||| |||||
Db 411 GACAGGACAGGTCTCAGAG 394

```

```

RESULT 12
US-08-154-019-29/c
; Sequence 29, Application US/08154019
; Patent No. 5633076
; GENERAL INFORMATION:
; APPLICANT: Deboer, Herman A.
; APPLICANT: Strijker, Rein
; APPLICANT: Heyneker, Herbert L.
; APPLICANT: Platenburg, Gerald
; APPLICANT: Lee, Sang He
; APPLICANT: Pieper, Frank
; APPLICANT: Krimpenfort, Paul J.A.
; TITLE OF INVENTION: Production of Recombinant Polypeptides
; TITLE OF INVENTION: by Bovine Species and Transgenic Methods
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/154,019
; FILING DATE: 16-NOV-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/461,333
; FILING DATE: 05-JUN-1995
; APPLICATION NUMBER: US 08/077,788

```

```

; FILING DATE: 15-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/895,956
; FILING DATE: 15-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/619,131
; FILING DATE: 27-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/444,745
; FILING DATE: 01-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Liebescheutz, Joe O.
; REGISTRATION NUMBER: 37,505
; REFERENCE/DOCKET NUMBER: 16994-003123
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 824 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-154-019-29

Query Match      82.2%; Score 14.8; DB 1; Length 824;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GACAGGGCAGCAGCAGAG 18
    ||||| |||| |||||
Db 417 GACAGGACAGGTCTCAGAG 400

RESULT 13
US-08-461-333-29/c
; Sequence 29, Application US/08461333
; Patent No. 5741957
; GENERAL INFORMATION:
; APPLICANT: Deboer, Herman A.
; APPLICANT: Strijker, Rein
; APPLICANT: Heyneker, Herbert L.
; APPLICANT: Platenburg, Gerald
; APPLICANT: Lee, Sang He
; APPLICANT: Pieper, Frank
; APPLICANT: Krimpenfort, Paul J.A.
; TITLE OF INVENTION: Production of Recombinant Polypeptides
; TITLE OF INVENTION: by Bovine Species and Transgenic Methods
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,333
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/077,788
; FILING DATE: 15-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/895,956
; FILING DATE: 15-JUN-1992
; PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: US 07/619,131  
FILING DATE: 27-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/444,745  
FILING DATE: 01-DEC-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Liebescheutz, Joe O.  
REGISTRATION NUMBER: 37,505  
REFERENCE/DOCKET NUMBER: 16994-003123  
TELEPHONE: 415-543-9600  
TELEFAX: 415-543-5043  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 824 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-461-333-29

Query Match 82.2%; Score 14.8; DB 1; Length 824;  
Best Local Similarity 88.9%; Pred. No. 1.9e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GACAAGGCGCAGACAGAG 18  
||||||| ||||| |||||  
Db 417 GACAAGGCGCAGACAGAG 400

RESULT 14  
US-08-461-167-29/c  
Sequence 29, Application US/08464167  
Patent No. 6013857  
GENERAL INFORMATION:  
APPLICANT: Deboer, Herman A.  
APPLICANT: Strijker, Rein  
APPLICANT: Heyneker, Herbert L.  
APPLICANT: Platenburg, Gerald  
APPLICANT: Lee, Sang He  
APPLICANT: Pieper, Frank  
APPLICANT: Krimpenfort, Paul J.A.  
TITLE OF INVENTION: Production of Recombinant Polypeptides  
TITLE OF INVENTION: by Bovine Species and Transgenic Methods  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: One Market Plaza, Steuart Tower, Suite 2000  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/464,167  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/077,788  
FILING DATE: 15-JUN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/895,956  
FILING DATE: 15-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/619,131  
FILING DATE: 27-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/444,745  
FILING DATE: 01-DEC-1989

ATTORNEY/AGENT INFORMATION:  
NAME: Liebescheutz, Joe O.  
REGISTRATION NUMBER: 37,505  
REFERENCE/DOCKET NUMBER: 16994-003124  
TELEPHONE: 415-543-9600  
TELEFAX: 415-543-5043  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 824 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-464-167-29

Query Match 82.2%; Score 14.8; DB 3; Length 824;  
Best Local Similarity 88.9%; Pred. No. 1.9e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GACAAGGCGCAGACAGAG 18  
||||||| ||||| |||||  
Db 417 GACAAGGCGCAGACAGAG 400

RESULT 15  
US-09-158-313-29/c  
Sequence 29, Application US/09158313  
Patent No. 6066725  
GENERAL INFORMATION:  
APPLICANT: Deboer, Herman A.  
APPLICANT: Strijker, Rein  
APPLICANT: Heyneker, Herbert L.  
APPLICANT: Platenburg, Gerald  
APPLICANT: Lee, Sang He  
APPLICANT: Pieper, Frank  
APPLICANT: Krimpenfort, Paul J.A.  
TITLE OF INVENTION: Production of Recombinant Polypeptides  
TITLE OF INVENTION: by Bovine Species and Transgenic Methods  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: One Market Plaza, Steuart Tower, Suite 2000  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/158,313  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/476,798  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/895,956  
FILING DATE: 15-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/619,131  
FILING DATE: 27-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/444,745  
FILING DATE: 01-DEC-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Liebescheutz, Joe O.  
REGISTRATION NUMBER: 37,505  
REFERENCE/DOCKET NUMBER: 16994-003125  
TELECOMMUNICATION INFORMATION:

```
;
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 824 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-158-313-29

Query Match      82.2%; Score 14.8; DB 3; Length 824;
Best Local Similarity 88.9%; Pred. NO. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GACAGGGCGAGCAGAG 18
        ||||||| |||| |||||
Db      417 GACAGGGCGAGCAGAG 400
```

```
Search completed: August 27, 2003, 19:42:01
Job time : 42.0385 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 27, 2003, 17:59:19 ; Search time 140.192 Seconds  
(without alignments)  
346.595 Million cell updates/sec

Title: US-10-085-612-1

Perfect score: 18

Sequence: 1 gacaaaggcagacagag 18

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	18	100.0	18	24	Human cytochrome P
2	16.4	91.1	22	24	Human promoter -39
3	16.4	91.1	29	20	PCR primer for Hum
4	16.4	91.1	1345	21	Nucleic acid seque
5	16.4	91.1	75236	25	Saccharopolyspora
6	16.4	91.1	123785	25	DNA sequence of hu
7	16.4	91.1	172637	24	Human voltage-acti
8	16.4	91.1	237961	24	Human Canion gene

9	16	88.9	1345	20	AA28296	Human CYP3A4 gene
10	16	88.9	1345	24	ABK68745	5'-flanking region
11	16	88.9	3881	21	AA281289	Human secreted pro
12	15.4	85.6	19	20	AA28287	Human CYP3A4 gene
13	15.4	85.6	19	24	ABK68748	Oligonucleotide #2
14	15.4	85.6	677	24	ABQ71128	Listeria monocytog
15	15.4	85.6	45546	20	AA23520	Human kidney amino
16	15.4	85.6	198161	24	ABK83564	Human CDNA differe
17	15	83.3	779	22	AAK86612	Human immune/haema
18	15	83.3	779	25	ABZ74371	Secreted protein g
19	15	83.3	779	25	ABZ67942	Human secreted pro
20	15	83.3	788	21	AA274381	Human secreted pro
21	15	83.3	788	25	ABZ73544	Secreted protein-e
22	15	83.3	788	25	ABZ67148	Human secreted pro
23	15	83.3	2224	24	ABQ78354	Nucleotide sequenc
24	15	83.3	3135	24	ABQ78353	Nucleotide sequenc
25	15	83.3	3276	24	AB199463	Mouse ischaemic co
26	14.8	82.2	18	24	AA245762	Human cytochrome p
27	14.8	82.2	21	24	AA236217	Human CYP3A4 promo
28	14.8	82.2	26	24	AA245776	Human promoter -39
29	14.8	82.2	32	20	AA28305	PCR primer for Hum
30	14.8	82.2	137	22	ABA50761	Human breast cell
31	14.8	82.2	137	22	ABA68729	Human foetal liver
32	14.8	82.2	137	22	ABA35692	Probe #14158 for g
33	14.8	82.2	137	22	AAK17072	Human brain expres
34	14.8	82.2	137	22	AAK42855	Human bone marrow
35	14.8	82.2	137	22	AAI23619	Probe #13552 for g
36	14.8	82.2	137	22	AAI48930	Probe #17616 used
37	14.8	82.2	137	22	AAI09234	Probe #9225 used t
38	14.8	82.2	137	23	ABS42487	Human liver single
39	14.8	82.2	137	24	ABS16909	Human genome-deriv
40	14.8	82.2	349	24	ABL79941	Human ovarian canc
41	14.8	82.2	352	24	ABN16225	Human ORFX polynu
42	14.8	82.2	374	22	AAF67509	Novel human polynu
43	14.8	82.2	381	22	AAF67510	Novel human polynu
44	14.8	82.2	429	22	ABA45639	Human breast cell
45	14.8	82.2	429	22	ABA56143	Human foetal liver

## ALIGNMENTS

RESULT 1

AA245760  
ID AAD45760 standard; DNR; 18 BP.

XX AAD45760;

XX 27-DEC-2002 (first entry)

XX Human cytochrome P450 (CYP) 3A4 polymorphic variant DNA fragment.

XX Human; cytochrome P450; CYP3A4; CYP3A5; glutathione S-transferase;  
KW GSTM1; breast cancer; therapy; chemotherapeutic agent; variant;  
KW drug-drug interaction; drug adverse effect; anti-cancer agent; SNP;  
KW enzyme; single nucleotide polymorphism; ds.

XX Homo sapiens.

OS Synthetic.

XX Key Location/Qualifiers  
FH variation replace (11, A)  
FT /tag= a

FT /standard\_name= 'Single nucleotide polymorphism (SNP)'

XX WO200268448-A1.

XX 06-SEP-2002.

XX 26-FEB-2002; 2002WO-US06135.

XX 26-FEB-2001; 2001US-271630P.

PA (DNAS-) DNA SCI INC.  
 PA (UYDU-) UNIV DUKE.  
 XX Guida M, Hall J, Petros WP, Vredenburg JU, Colvin OM, Marks JR;  
 XX WPI; 2002-691652/74.  
 XX  
 XX New nucleic acid molecule useful for identifying polymorphisms  
 PT associated with CYP3A4, CYP3A5 or GSTM1 substrates which influence  
 PT breast cancer treatments, comprises at least one base variation from  
 PT human CYP3A4 or CYP3A5 sequence -  
 XX Claim 1; Page 12; 41pp; English.  
 XX  
 XX The invention relates to a nucleic acid molecule comprising at least one  
 CC base variation from human cytochrome P450 (CYP) 3A4 or CYP3A5 sequence.  
 CC Nucleic acid molecules of the invention are useful for identifying  
 CC polymorphisms associated with CYP3A4, CYP3A5 or glutathione S-transferase  
 CC (GST) M1 substrates which influence breast cancer treatments. They are  
 CC also useful in diagnostic purposes to identify individuals having a  
 CC polymorphic genotype which influence the outcome of breast cancer  
 CC treatments and the selection of chemotherapeutic agents used to treat  
 CC breast cancer. The polymorphisms detected are used to screen altered  
 CC metabolism of CYP3A4, CYP3A5 or GSTM1 substrates, potential drug-drug  
 CC interactions, drug adverse effects, likelihood of successful clinical  
 CC outcome following treatment with anti-cancer agents such as cisplatin,  
 CC cyclophosphamide and/or BCNU. The present sequence is human CYP3A4  
 CC polymorphic variant DNA fragment.  
 XX Sequence 18 BP; 7 A; 3 C; 8 G; 0 U; 0 other;  
 SQ  
 Query Match 100.0%; Score 18; DB 24; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GACAGGGCGAGCAGACAG 18  
 Db 1 GACAGGGCGAGCAGACAG 18  
 RESULT 2  
 AAD45777  
 ID AAD45777 standard; DNA; 22 BP.  
 XX  
 XX AAD45777;  
 AC  
 DT 27-DEC-2002 (first entry)  
 DE Human promoter -392 CYP4503A4 specific probe #2.  
 XX  
 KW Human; cytochrome P450; CYP3A4; CYP3A5; glutathione S-transferase;  
 KW polymorphism; GSTM1; breast cancer; therapy; chemotherapeutic agent;  
 KW drug-drug interaction; drug adverse effect; anti-cancer agent;  
 KW enzyme; probe; promoter; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200268448-A1.  
 XX  
 PD 06-SEP-2002.  
 XX  
 PF 26-FEB-2002; 2002WO-US06135.  
 XX  
 PR 26-FEB-2001; 2001US-271630P.  
 XX  
 PA (DNAS-) DNA SCI INC.  
 PA (UYDU-) UNIV DUKE.  
 XX  
 XX Guida M, Hall J, Petros WP, Vredenburg JU, Colvin OM, Marks JR;  
 XX WPI; 2002-691652/74.  
 XX  
 XX New nucleic acid molecule useful for identifying polymorphisms

PT associated with CYP3A4, CYP3A5 or GSTM1 substrates which influence  
 PT breast cancer treatments, comprises at least one base variation from  
 PT human CYP3A4 or CYP3A5 sequence -  
 XX Example 2; Page 26; 41pp; English.  
 XX  
 XX The invention relates to a nucleic acid molecule comprising at least one  
 CC base variation from human cytochrome P450 (CYP) 3A4 or CYP3A5 sequence.  
 CC Nucleic acid molecules of the invention are useful for identifying  
 CC polymorphisms associated with CYP3A4, CYP3A5 or glutathione S-transferase  
 CC (GST) M1 substrates which influence breast cancer treatments. They are  
 CC also useful in diagnostic purposes to identify individuals having a  
 CC polymorphic genotype which influence the outcome of breast cancer  
 CC treatments and the selection of chemotherapeutic agents used to treat  
 CC breast cancer. The polymorphisms detected are used to screen altered  
 CC metabolism of CYP3A4, CYP3A5 or GSTM1 substrates, potential drug-drug  
 CC interactions, drug adverse effects, likelihood of successful clinical  
 CC outcome following treatment with anti-cancer agents such as cisplatin,  
 CC cyclophosphamide and/or BCNU. The present sequence is human promoter  
 CC -392 CYP4503A4 specific probe.  
 XX Sequence 22 BP; 8 A; 3 C; 11 G; 0 U; 0 other;  
 SQ  
 Query Match 91.1%; Score 16.4; DB 24; Length 22;  
 Best Local Similarity 94.4%; Pred. No. 1.8e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GACAGGGCGAGCAGACAG 18  
 Db 1 GACAGGGCGAGCAGACAG 18  
 RESULT 3  
 AAX28306/c  
 ID AAX28306 standard; DNA; 29 BP.  
 XX  
 XX AAX28306;  
 AC  
 DT 17-JUN-1999 (first entry)  
 DE PCR primer for Human CYP3A4 gene promoter.  
 XX  
 KW CYP3A4 gene polymorphism; polymorphic locus; human; altered metabolism;  
 KW CYP3A4 substrate; drug-drug interaction identification; toxin exposure;  
 KW genetic linkage detection; phenotypic variation; promoter; PCR primer;  
 KW ss.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 PN WO913106-A1.  
 XX  
 PD 18-MAR-1999.  
 XX  
 PF 02-SEP-1998; 98WO-US18158.  
 XX  
 PR 10-SEP-1997; 97US-0058612.  
 XX  
 PA (AXIS-) AXIS PHARM INC.  
 XX  
 PI Guida M, Lichter JB;  
 XX  
 DR WPI; 1999-215070/18.  
 XX  
 XX New isolated CYP3A4 polymorphic sequences  
 PT  
 PS Example; Page 15; 40pp; English.  
 XX  
 CC This sequence represents a PCR primer for the human CYP3A4 gene promoter.  
 CC The invention relates to a CYP3A4 sequence polymorphism,  
 CC which is part of a non-naturally occurring chromosome. Nucleic acids  
 CC comprising the CYP3A4 polymorphic sequences can be used to screen  
 CC patients for altered metabolism for CYP3A4 substrates, potential

CC drug-drug interactions, and adverse/side effects as well as diseases that  
 CC result from environmental or occupational exposure to toxins. They can  
 CC also be used to establish animal, cell culture and in vitro cell-free  
 CC models for drug metabolism. Polymorphic CYP3A4 gene sequences can be used  
 CC for expression studies to determine the effect of promoter and/or intron  
 CC sequence variations on mRNA expression and stability. The polymorphisms  
 CC are also used as single nucleotide polymorphisms to detect genetic  
 CC linkage to phenotypic variation in activity and expression of CYP3A4. The  
 CC nucleic acids can also be used to generate genetically modified non-human  
 CC animals or site specific gene modifications in cell lines.

XX Sequence 29 BP; 3 A; 12 C; 3 G; 11 T; 0 other;

Query Match 91.1%; Score 16.4; DB 20; Length 29;

Best Local Similarity 94.4%; Pred. No. 1.8e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACAAGGCGCAGGAGAG 18  
 |||||  
 Db 24 GACAAGGCGCAGGAGAG 7

RESULT 4  
 AAZ57020  
 ID AAZ57020 standard; DNA; 1345 BP.

AC AAZ57020;

DT 19-MAY-2000 (first entry)

XX Nucleic acid sequence of CYP3A4 variant CYP3A4-V.

DE CYP3A4; cytochrome P450; variant; prostate cancer; leukemia;

KW epipodophyllotoxin; human; anticancer; ds.

OS Homo sapiens.

PN W0200009752-A1.

PD 24-FEB-2000.

PF 11-AUG-1999; 99WO-US18266.

PR 14-AUG-1998; 98US-0096586.

PA (TYPE-) UNIV PENNSYLVANIA.

FA (CHIL-) CHILDRENS HOSPITAL PHILADELPHIA.

PI Rebbeck TR, Felix CA;

DR WPI; 2000-224371/19.

XX Nucleic acid sequence encoding a variant of CYP3A4, involved in the  
 PT metabolism of aflatoxin B1 and drugs, useful for identifying patients  
 PT with increased risk of developing or having prostate cancer

PS Claim 2; Page 33; 36pp; English.

XX The invention provides a nucleic acid sequence encoding a variant of  
 CC CYP3A4, a member of cytochrome P450 supergene family. The CYP3A4 variant  
 CC is useful for identifying patients with heightened risk of developing or  
 CC having prostate cancer and at risk for developing treatment-related  
 CC leukemia upon administration of an epipodophyllotoxin. By identifying the  
 CC CYP3A4 variant, a more effective anticancer treatment regimen can be  
 CC selected. The present sequence represents the nucleic acid sequence of  
 CC the CYP3A4 variant CYP3A4-V. This variant comprises an A to G transition  
 CC that alters the 10 basepair nifedipine-specific element located at -287  
 CC to -296 base pairs from the transcription start site of CYP3A4.

XX Sequence 1345 BP; 352 A; 314 C; 357 G; 322 T; 0 other;

Query Match 91.1%; Score 16.4; DB 21; Length 1345;

Best Local Similarity 94.4%; Pred. No. 2.2e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACAAGGCGCAGGAGAG 18  
 |||||  
 Db 806 GACAAGGCGCAGGAGAG 823

RESULT 5  
 ABV75557  
 ID ABV75557 standard; DNA; 75236 BP.

XX AC ABV75557;

XX 22-JAN-2003 (first entry)

XX Saccharopolyspora butenyl-spinosyn biosynthetic gene cluster 1.

KW Butenyl; biosynthetic enzyme; PKS; polyketide synthetase; macrolide;  
 KW metabolite; spinosyn; gene; ds.

OS Saccharopolyspora sp.

XX Key Location/Qualifiers

PH 1..13035

FT /\*tag= a

FT /product= "busA"

FT 13059..19508

FT /\*tag= b

FT /product= "busB"

FT /note= "No start codon given"

FT 19553..29056

FT /\*tag= c

FT /product= "busC"

FT 29092..43893

FT /\*tag= d

FT /product= "busD"

FT 43945..60639

FT /\*tag= e

FT /product= "busE"

FT 62090..63940

FT /\*tag= f

FT /product= "ORF RI"

FT /transl\_except= (pos:62879..62881,aa:Xaa)

FT /transl\_except= (pos:62939..62941,aa:Xaa)

FT /note= "Xaa=Unknown. No start codon given"

FT complement (65226..66602)

FT /\*tag= g

FT /product= "ORF RII"

FT /note= "No start codon given"

FT complement (68759..69676)

FT /\*tag= h

FT /product= "ORF RIII"

FT /note= "No start codon given"

XX W0200279477-A2.

XX PD 10-OCT-2002.

XX 28-MAR-2002; 2002WO-US09968.

XX 30-MAR-2001; 2001US-280175P.

XX (DOWC ) DOW AGROSCIENCES LLC.

XX Hahn DR, Jackson JD, Bullard BS, Gustafson GD, Waldron C;

PI Mitchell JC;

XX WPI; 2003-058434/05.

DR P-ESDS; ABP57678, ABP57679, ABP57680, ABP57681, ABP57682, ABP57705,

DR ABP57706, ABP57707.

XX New butenyl-spinosyn biosynthetic genes, useful for increasing the

PT production of butenyl-spinosyn insecticidal macroides, or for changing

PT the metabolites or products produced by spinosyn-producing  
 PT microorganisms  
 XX  
 PS Claim 2; Page 57-99; 218pp; English.  
 XX  
 CC The invention relates to a novel DNA molecule comprising a DNA sequence  
 CC that encodes a butenyl-spinosyn biosynthetic enzyme, a butenyl-spinosyn  
 CC PKS (polyketide synthetase) domain, or a spinosyn PKS module. The  
 CC butenyl-spinosyn biosynthetic genes are useful for increasing the  
 CC production of butenyl-spinosyn insecticidal macrolides. The genes are  
 CC also useful for changing the metabolites or products produced by  
 CC spinosyn-producing microorganisms. The present sequence represents a DNA  
 CC molecule encoding butenyl-spinosyn biosynthetic enzymes.  
 XX  
 SQ Sequence 75236 BP; 10935 A; 21693 C; 28185 G; 14421 T; 2 other;  
 Query Match 91.1%; Score 16.4; DB 25; Length 75236;  
 Best Local Similarity 94.4%; Pred. No. 2.7e-02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GACAAGGGCAGGACAGAG 18  
 DB 63954 GACAAGGGCAGGACGAG 63971  
 RESULT 6  
 ABX77171  
 ID ABX77171 standard; DNA; 123785 BP.  
 XX  
 AC ABX77171;  
 DT 25-APR-2003 (first entry)  
 DE DNA sequence of human BAC clone RP11-757A13.  
 XX  
 KW Human; ss; transgenic; drug metabolism; behaviour; gene; mouse;  
 KW pharmacokinetic assay; pharmacodynamic assay; toxicology; serum albumin;  
 KW alpha-acidic glycoprotein; CYP; multidrug resistance protein; MRP;  
 KW uridine diphosphoglucuronosyl transferase; UGT; cytochrome P450.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200283897-A1.  
 XX  
 PD 24-OCT-2002.  
 XX  
 XX 18-APR-2002; 2002WO-AU000485.  
 XX  
 XX 18-APR-2001; 2001AU-0004467.  
 XX  
 PA (GENE-) GENE STREAM PTY LTD.  
 XX  
 PI Daly JM;  
 XX  
 DR WPI; 2003-093021/08.  
 XX  
 XX New transgenic non-human animal expressing a foreign polypeptide  
 PT associated with drug behavior and/or metabolism, useful for studying  
 PT the behavior and/or metabolism of a drug in other animals -  
 XX  
 PS Example 2A; Page 229-295; 408pp; English.  
 XX  
 CC This invention relates to a transgenic non-human animal which may be  
 CC used for assessing the behaviour and/or metabolism of a drug in another  
 CC animal and which expresses a foreign polypeptide associated with drug  
 CC behaviour and/or metabolism. The invention also comprises a nucleic acid  
 CC construct for use in producing the above transgenic non-human animal  
 CC and a method of assessing the metabolism and/or behavior of a drug in  
 CC an animal of interest, comprising administering a test agent to the  
 CC transgenic animal and conducting analytical tests to determine drug  
 CC metabolism and/or behaviour. The transgenic animal is useful in studying  
 CC drug metabolism and/or behaviour in other animals. The nucleic acid  
 CC construct is useful in producing the above transgenic animal and the

CC methods are used for producing, breeding and using transgenic animals  
 CC for pharmacological (e.g. pharmacokinetic or pharmacodynamic assays)  
 CC and/or toxicological studies. Nucleic acid sequences used within  
 CC the invention are serum albumin; alpha-acidic glycoprotein; cytochrome  
 CC P450 (CYP); uridine diphosphoglucuronosyl transferase (UGT); multidrug  
 CC resistance proteins and (MRP's). The present sequence represents a  
 CC DNA sequence used to create a transgenic animal within the scope of the  
 CC invention.  
 XX  
 SQ Sequence 123785 BP; 34793 A; 24793 C; 26537 G; 37655 T; 7 other;  
 Query Match 91.1%; Score 16.4; DB 25; Length 123785;  
 Best Local Similarity 94.4%; Pred. No. 2.7e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GACAAGGGCAGGACAGAG 18  
 DB 34911 GACAAGGGCAGGACAGAG 34928  
 RESULT 7  
 ABN83124  
 ID ABN83124 standard; DNA; 172637 BP.  
 XX  
 AC ABN83124;  
 DT 09-SEP-2002 (first entry)  
 DE Human voltage-activated ion channel transporter protein gene.  
 XX  
 KW Human; transporter protein; voltage-activated ion channel; gene therapy;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key  
 FT variation  
 FT replace (1543,C)  
 FT /tag= a  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace (1546,G)  
 FT /tag= b  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace (2514..2515,ATC/ACC)  
 FT /tag= c  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT 3000..170072  
 FT /tag= d  
 FT /product= "Transporter protein"  
 FT /note= "Contains 11 introns"  
 FT 3000..3107  
 FT /tag= e  
 FT /number= 1  
 FT 3108..6761  
 FT /tag= f  
 FT /number= 1  
 FT /tag= g  
 FT replace (3387,T)  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace (3638,G)  
 FT /tag= h  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace (3817,G)  
 FT /tag= i  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT replace (8649,T)  
 FT /tag= j  
 FT /standard\_name= "Single nucleotide polymorphism"  
 FT 6762..6944  
 FT /tag= k  
 FT /number= 2  
 FT 6945..23449  
 FT /tag= l  
 FT /number= 2

FT	variation	replace (8145,A)	/*tag= m	/standard_name= "Single nucleotide polymorphism"	FT	/*tag= ak	/standard_name= "Single nucleotide polymorphism"
FT	variation	replace (8999,T)	/*tag= n	replace (8999,T)	FT	replace (30649,G)	/*tag= al
FT	variation	replace (9044,C)	/*tag= o	replace (9044,C)	FT	replace (31479,G)	/standard_name= "Single nucleotide polymorphism"
FT	variation	replace (9990,A)	/*tag= p	replace (9990,A)	FT	replace (31835,A)	/*tag= am
FT	variation	replace (10712,T)	/*tag= q	replace (10712,T)	FT	replace (35034,A)	/standard_name= "Single nucleotide polymorphism"
FT	variation	replace (11855,C)	/*tag= r	replace (11855,C)	FT	replace (35183,G)	/*tag= au
FT	variation	replace (12177,G)	/*tag= s	replace (12177,G)	FT	replace (47020,T)	/standard_name= "Single nucleotide polymorphism"
FT	variation	replace (12328,T)	/*tag= t	replace (12328,T)	FT	replace (47264,C)	/*tag= ag
FT	variation	replace (13406,A)	/*tag= u	replace (13406,A)	FT	replace (52397,T)	/standard_name= "Single nucleotide polymorphism"
FT	variation	replace (16751,C)	/*tag= v	replace (16751,C)	FT	replace (52397,T)	/*tag= ar
FT	variation	replace (20631,G)	/*tag= w	replace (20631,G)	FT	replace (56497..56651)	/standard_name= "Single nucleotide polymorphism"
FT	variation	replace (20587,G)	/*tag= x	replace (20587,G)	FT	/*tag= au	/*tag= 6
FT	variation	replace (20793,A)	/*tag= y	replace (20793,A)	FT	56652..109657	/*tag= av
FT	variation	replace (20935,G)	/*tag= z	replace (20935,G)	FT	/*tag= 6	/*tag= 6
FT	exon	23450..23533	/*tag= aa	23450..23533	FT	56652..109657	/*tag= av
FT	intron	23534..25047	/*tag= ab	23534..25047	FT	/*tag= 6	/*tag= 6
FT	exon	25048..25187	/*tag= ac	25048..25187	FT	replace (56912,G)	/*tag= aw
FT	intron	25188..25275	/*tag= ad	25188..25275	FT	replace (56912,G)	/*tag= ax
FT	exon	25276..25404	/*tag= ae	25276..25404	FT	replace (60709,A)	/standard_name= "Single nucleotide polymorphism"
FT	intron	25405..56496	/*tag= af	25405..56496	FT	replace (60709,A)	/*tag= ay
FT	variation	replace (26154,C)	/*tag= ag	replace (26154,C)	FT	replace (67297,T)	/standard_name= "Single nucleotide polymorphism"
FT	variation	replace (27330,C)	/*tag= ah	replace (27330,C)	FT	replace (67297,T)	/*tag= az
FT	variation	replace (27602,T)	/*tag= ai	replace (27602,T)	FT	replace (70472,T)	/standard_name= "Single nucleotide polymorphism"
FT	variation	replace (27615..27616,G)	/*tag= aj	replace (27615..27616,G)	FT	replace (70472,T)	/*tag= ba
FT	variation	replace (30002,G)	/*tag= ak	replace (30002,G)	FT	replace (70899,G)	/standard_name= "Single nucleotide polymorphism"
FT	variation	replace (30002,G)	/*tag= al	replace (30002,G)	FT	replace (70899,G)	/*tag= bb
FT	variation	replace (30002,G)	/*tag= am	replace (30002,G)	FT	replace (71170,T)	/standard_name= "Single nucleotide polymorphism"
FT	variation	replace (30002,G)	/*tag= an	replace (30002,G)	FT	replace (72558,G)	/*tag= bc
FT	variation	replace (30002,G)	/*tag= ao	replace (30002,G)	FT	replace (72558,G)	/*tag= bd
FT	variation	replace (30002,G)	/*tag= ap	replace (30002,G)	FT	replace (73323,G)	/standard_name= "Single nucleotide polymorphism"
FT	variation	replace (30002,G)	/*tag= aq	replace (30002,G)	FT	replace (73323,G)	/*tag= be
FT	variation	replace (30002,G)	/*tag= ar	replace (30002,G)	FT	replace (73438,A)	/standard_name= "Single nucleotide polymorphism"
FT	variation	replace (30002,G)	/*tag= as	replace (30002,G)	FT	replace (73438,A)	/*tag= bf
FT	variation	replace (30002,G)	/*tag= at	replace (30002,G)	FT	replace (73818,G)	/standard_name= "Single nucleotide polymorphism"
FT	variation	replace (30002,G)	/*tag= au	replace (30002,G)	FT	replace (73818,G)	/*tag= bg
FT	variation	replace (30002,G)	/*tag= av	replace (30002,G)	FT	replace (74035,C)	/standard_name= "Single nucleotide polymorphism"
FT	variation	replace (30002,G)	/*tag= aw	replace (30002,G)	FT	replace (74035,C)	/*tag= bh
FT	variation	replace (30002,G)	/*tag= ax	replace (30002,G)	FT	replace (74832,C)	/standard_name= "Single nucleotide polymorphism"
FT	variation	replace (30002,G)	/*tag= ay	replace (30002,G)	FT	replace (74832,C)	/*tag= bi
FT	variation	replace (30002,G)	/*tag= az	replace (30002,G)	FT	replace (74832,C)	/*tag= bi



```

FT FT /standard_name= "Single nucleotide polymorphism"
FT FT replace (75509,T)
FT FT /*tag= b_j

Query Match 91.1%; Score 16.4; DB 24; Length 172637;
Best Local Similarity 94.4%; Pred. No. 2.7e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAGGGGAGGACAGAG 18
Db 95503 GACAGGGGAGGACAGAG 95520

RESULT 8
ABQ80552
ID ABQ80552 standard; DNA; 237961 BP.
XX AC
XX ABQ80552;
DT 08-NOV-2002 (first entry)
XX DE Human Canion gene fragment #2.
XX KW Human; Canion; neuroleptic; hypotensive; anticonvulsant; analgesic;
KW antiarrhythmic; antianginal; cardiant; antitmanic; antidepressant;
KW gene therapy; schizophrenia; voltage-gated ion channel; bipolar disorder;
KW central nervous system disorder; cardiovascular disorder; hypertension;
KW pain; epilepsy; chromosome 13q; gene; ds.
XX OS Homo sapiens.
XX FH Key
XX FH Location/Qualifiers
FT exon 43726..43868
FT /*tag= a
FT /*number= 8
FT exon 43998..44102
FT /*tag= b
FT /*number= 9
FT misc_difference 51090
FT /*tag= c
FT /*note= "Biallelic marker: 99-79335-60/A7"
FT exon 52093..52179
FT /*tag= d
FT /*number= 10
FT misc_difference 61293
FT /*tag= e
FT /*note= "Biallelic marker: 99-79336-369/A8"
FT exon 77568..77699
FT /*tag= f
FT /*number= 11
FT misc_difference 80602
FT /*tag= g
FT /*note= "Biallelic marker: 99-79338-332/A9"
FT exon 98226..98393
FT /*tag= h
FT /*number= 12
FT misc_difference 100485
FT /*tag= i
FT /*note= "Biallelic marker: 99-79314-201/A10"
FT misc_difference 100509
FT /*tag= j
FT /*note= "Biallelic marker: 99-79314-225/A11"
FT exon 106567..106758
FT /*tag= k
FT /*number= 13
FT misc_difference 106725
FT /*tag= l
FT /*note= "Biallelic marker: 99-79316-158/A12"
FT exon 144109..144246
FT /*tag= m
FT /*number= 14
FT exon 159797..159868
FT /*tag= n

/number= 15
misc_difference 166087
/*tag= o
/*note= "Biallelic marker: 99-79322-224/A13"
misc_difference 166336
/*tag= p
/*note= "Biallelic marker: 99-79322-473/A14"
exon 191292..191428
/*tag= q
/*number= 16
exon 192967..193108
/*tag= r
/*number= 17
exon 211540..211613
/*tag= s
/*number= 18
exon 225006..225107
/*tag= t
/*number= 19
exon 225544..225613
/*tag= u
/*number= 20
exon 228450..228541
/*tag= v
/*number= 21
exon 228630..228752
/*tag= x
/*number= 22
exon 231289..231345
/*tag= y
/*number= 23
exon 231589..231709
/*tag= z
/*number= 24
exon 231813..231944
/*tag= aa
/*number= 25
exon 232900..233067
/*tag= ab
/*number= 26
exon 235355..235459
/*tag= ac
/*number= 27
misc_difference 235894
/*tag= ad
/*note= "Biallelic marker: 99-79306-182/A15"
WO200246404-A2.
13-JUN-2002.
04-DEC-2001; 2001WO-IB02798.
05-DEC-2000; 2000US-251317P.
(GEST ) GENSET.
Cohen D, Chumakov I, Simon A, Abderrahim H;
WPI; 2002-619018/66.
Novel schizophrenia-related voltage-gated ion channel polypeptide and
polynucleotide useful for identifying modulators and for diagnosing,
treating schizophrenia, bipolar disorder or central nervous system
disorders
Claim 1; Page 183-246; 272pp; English.
The present sequence is a gene fragment for human Canion, a novel
schizophrenia-related voltage-gated ion channel protein. Canion and its
coding sequence are useful in the treatment of schizophrenia, bipolar
disorder, or other central nervous system (CNS) conditions, as well as
other conditions such as heart conditions and hypertension. Compounds

```

CC that block Canion channels are useful to treat a number of diseases or  
 CC conditions, preferably schizophrenia or bipolar disorder, and also  
 CC including pain disorders, epilepsy and various cardiovascular disorders  
 CC such as heart arrhythmias, angina and hypertension. The Canion gene is  
 CC located on chromosome 13q. The present sequence comprises a genomic  
 CC sequence of Canion, comprising exons 8 to 27.

XX  
 SQ Sequence 237961 BP; 74252 A; 43806 C; 46176 G; 72420 T; 1307 other;

Query Match 91.1%; Score 16.4; DB 24; Length 237961;  
 Best Local Similarity 94.4%; Pred. No. 2.8e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAAGGGCAGGACAGAG 18  
 |||||  
 Db 29559 GACAAGGGCAGGACAGAG 29576

RESULT 9  
 AAX28296  
 ID AAX28296 standard; DNA; 1345 BP.

XX AC AAX28296;

XX DT 17-JUN-1999 (first entry)

XX DE Human CYP3A4 gene promoter.

XX CYP3A4 gene polymorphism; polymorphic locus; human; altered metabolism;  
 KW CYP3A4 substrate; drug-drug interaction identification; toxin exposure;  
 KW genetic linkage detection; phenotypic variation; promoter; ss.

XX OS Homo sapiens.

XX PN W09913106-A1.

XX PD 18-MAR-1999.

XX PF 02-SEP-1998; 98WO-US18158.

XX PR 10-SEP-1997; 97US-0058612.

XX PA (AXYS-) AXYS PHARM INC.

XX PI Guida M, Lichter JB;

XX DR WPI; 1999-215070/18.

XX PT New isolated CYP3A4 polymorphic sequences

XX PS Disclosure; Page 29; 40pp; English.

XX This sequence represents the human CYP3A4 gene promoter.  
 CC The invention relates to a CYP3A4 sequence polymorphism,  
 CC which is part of a non-naturally occurring chromosome. Nucleic acids  
 CC comprising the CYP3A4 polymorphic sequences can be used to screen  
 CC patients for altered metabolism for CYP3A4 substrates, potential  
 CC drug-drug interactions, and adverse/side effects as well as diseases that  
 CC result from environmental or occupational exposure to toxins. They can  
 CC also be used to establish animal, cell culture and in vitro cell-free  
 CC models for drug metabolism. Polymorphic CYP3A4 gene sequences can be used  
 CC for expression studies to determine the effect of promoter and/or intron  
 CC sequence variations on mRNA expression and stability. The polymorphisms  
 CC are also used as single nucleotide polymorphisms to detect genetic  
 CC linkage to phenotypic variation in activity and expression of CYP3A4. The  
 CC nucleic acids can also be used to generate genetically modified non-human  
 CC animals or site specific gene modifications in cell lines.

XX SQ Sequence 1345 BP; 352 A; 314 C; 356 G; 322 T; 1 other;

Query Match 88.9%; Score 16; DB 20; Length 1345;  
 Best Local Similarity 88.9%; Pred. No. 3.4e+02;  
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAAGGGCAGGACAGAG 18  
 |||||  
 Db 806 GACAAGGGCAGGACAGAG 823

RESULT 10  
 ABK68745  
 ID ABK68745 standard; DNA; 1345 BP.

XX AC ABK68745;

XX DT 02-JUL-2002 (first entry)

XX DE 5'-flanking region of human CYP3A4 gene.

XX KW Human; single nucleotide polymorphism; SNP; cytochrome p450; CYP;  
 KW CYP3A4; ds.

XX OS Homo sapiens.

XX PN W0200218641-A2.

XX PD 07-MAR-2002.

XX PF 30-AUG-2001; 2001WO-IB01580.

XX PR 30-AUG-2000; 2000GB-0021286.

XX PA (GEMI-) GEMINI GENOMICS PLC.

XX PI Risinger C, Andersson MK, Lewander T, Olaisson E;

XX DR WPI; 2002-351712/38.

XX PT Novel primer pairs and sequence determination oligonucleotides useful  
 PT for amplifying and detecting novel single nucleotide polymorphisms in  
 PT the 5'-flanking regions of cytochrome p450 (CYP)3A4 and CYP2C9 genes  
 PT respectively

XX PS Claim 1; Fig 1; 47pp; English.

XX The present invention relates to PCR primer pairs for amplifying  
 CC and sequence determination oligonucleotides for detecting single  
 CC nucleotide polymorphisms (SNPs) in the 5'-flanking regions of human  
 CC cytochrome p450 (CYP) genes encoding CYP3A4 or CYP2C9. The SNPs  
 CC correspond to position 461 of a defined 1345 base pair sequence for  
 CC CYP3A4 or position 957, 1049, 1164, 1326, 1661 and 1662 of a 2438 base  
 CC pair sequence for CYP2C9. The PCR primers are useful for amplifying  
 CC the CYP sequences and the oligonucleotides are useful for detecting  
 CC SNPs in the 5'-flanking regions of the CYP3A4 or CYP2C9 genes. The  
 CC present sequence represents the 5'-flanking region of the human CYP3A4  
 CC gene.

XX SQ Sequence 1345 BP; 352 A; 313 C; 356 G; 322 T; 2 other;

Query Match 88.9%; Score 16; DB 24; Length 1345;  
 Best Local Similarity 88.9%; Pred. No. 3.4e+02;  
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAAGGGCAGGACAGAG 18  
 |||||  
 Db 806 GACAAGGGCAGGACAGAG 823

RESULT 11  
 AAA61269/c  
 ID AAA61269 standard; DNA; 3881 BP.

XX AC AAA61269;

XX DT 18-OCT-2000 (first entry)

DE Human secreted protein gene 10 clone HDPGP94.  
 XX Human; secreted protein; fusion protein; gene therapy;  
 KW protein therapy; diagnosis; tissue; cancer; tumour; AIDS;  
 KW autoimmune disorder; allergy; cardiovascular; viral; bacterial;  
 KW fungal infection; immunosuppressive; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200029422-A1.  
 XX  
 XX 25-MAY-2000.  
 XX  
 XX 09-NOV-1999; 99WO-US26409.  
 XX  
 XX 12-NOV-1998; 98US-0108207.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Ni J, Ruben SM, Rosen CA, Ebner R, Florence KA, Young PE;  
 PI Birse CE, Carter KC, Komatsoulis G;  
 XX  
 XX WPI; 2000-387729/33.  
 XX  
 XX Novel human secreted proteins useful for diagnosing, preventing,  
 PT treating and ameliorating a medical condition e.g. cardiovascular  
 PT disease -  
 XX  
 XX Claim 1; Page 233-234; 295pp; English.  
 XX  
 XX The present sequence represents a nucleic acid molecule which encodes a  
 CC secreted human protein. The gene number and the clone it was derived  
 CC from are given in the descriptor line.  
 CC The invention relates to 31 novel genes and their fragments (nucleic  
 CC acid sequences: AAA61260-A61293; amino acid sequences AAB12301-B12371)  
 CC which are useful for preventing, treating or ameliorating medical  
 CC conditions e.g. by protein or gene therapy. Also pathological conditions  
 CC can be diagnosed by determining the amount of the new polypeptides in a  
 CC sample or by determining the presence of mutations in the new  
 CC polynucleotides. Specific uses are described for each of the 31  
 CC polynucleotides, based on which tissues they are most highly expressed in  
 CC and include products for the diagnosis or treatment of cancer, tumours,  
 CC AIDS, autoimmune disorders, allergy, cardiovascular disorders, viral,  
 CC bacterial and fungal infection. The genes are used to generate fusion  
 CC proteins by linking to the gene a human immunoglobulin portion (AAA61251)  
 CC for increasing stability of the fused protein as compared to the  
 CC secreted protein only.  
 XX  
 XX Sequence 3881 BP; 1043 A; 610 C; 716 G; 1512 T; 0 other;  
 SQ  
 Query Match 88.9%; Score 16; DB 21; Length 3881;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 3 CAAGGGCAGGACAGAG 18  
 Db 2592 CAAGGGCAGGACAGAG 2577  
 |||||  
 RESULT 12  
 AAX28287  
 ID AAX28287 standard; DNA; 19 BP.  
 XX  
 XX AAX28287;  
 AC  
 XX 17-JUN-1999 (first entry)  
 DT  
 XX Human CYP3A4 gene polymorphism #1.  
 DE  
 XX CYP3A4 gene polymorphism; polymorphic locus; human; altered metabolism;  
 KW CYP3A4 substrate; drug-drug interaction identification; toxin exposure;  
 KW genetic linkage detection; phenotypic variation; ss.  
 XX

OS Homo sapiens.  
 XX  
 XX WO9913106-A1.  
 XX  
 XX 18-MAR-1999.  
 PD  
 XX 02-SEP-1998; 98WO-US18158.  
 PF  
 XX 10-SEP-1997; 97US-0058612.  
 PR  
 XX (AXYS-) AXYS PHARM INC.  
 XX  
 XX Guida M, Lichter JB;  
 PI  
 XX WPI; 1999-215070/18.  
 DR  
 XX New isolated CYP3A4 polymorphic sequences  
 XX  
 XX Claim 2; Page 35; 40pp; English.  
 XX  
 XX This sequence represents a CYP3A4 sequence polymorphism of the invention,  
 CC which is part of a non-naturally occurring chromosome. Nucleic acids  
 CC comprising the CYP3A4 polymorphic sequences can be used to screen  
 CC patients for altered metabolism for CYP3A4 substrates, potential  
 CC drug-drug interactions, and adverse/side effects as well as diseases that  
 CC result from environmental or occupational exposure to toxins. They can  
 CC also be used to establish animal, cell culture and in vitro cell-free  
 CC models for drug metabolism. Polymorphic CYP3A4 gene sequences can be used  
 CC for expression studies to determine the effect of promoter and/or intron  
 CC sequence variations on mRNA expression and stability. The polymorphisms  
 CC are also used as single nucleotide polymorphisms to detect genetic  
 CC linkage to phenotypic variation in activity and expression of CYP3A4. The  
 CC nucleic acids can also be used to generate genetically modified non-human  
 CC animals or site specific gene modifications in cell lines.  
 XX  
 XX Sequence 19 BP; 7 A; 3 C; 9 G; 0 U; 0 other;  
 SQ  
 Query Match 85.6%; Score 15.4; DB 20; Length 19;  
 Best Local Similarity 94.1%; Pred. No. 5.3e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 2 ACAAGGGCAGGACAGAG 18  
 Db 1 ACAAGGGCAGGACAGAG 17  
 |||||  
 RESULT 13  
 ABK68748  
 ID ABK68748 standard; DNA; 19 BP.  
 XX  
 XX ABK68748;  
 AC  
 XX 02-JUL-2002 (first entry)  
 DT  
 XX Oligonucleotide #2 for detecting polymorphism in CYP3A4 gene.  
 DE Human; single nucleotide polymorphism; SNP; cytochrome p450; CYP;  
 XX CYP3A4; ss.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200218641-A2.  
 PN  
 XX 07-MAR-2002.  
 PD  
 XX 30-AUG-2001; 2001WO-IB01580.  
 PF  
 XX 30-AUG-2000; 2000GB-0021286.  
 PR  
 XX (GEMI-) GEMINI GENOMICS PLC.  
 XX  
 XX Risinger C, Andersson MK, Lewander T, Olaiasson E;  
 PI

DR WPI; 2002-351712/38.

XX Novel primer pairs and sequence determination oligonucleotides useful

PT for amplifying and detecting novel single nucleotide polymorphisms in

PT the 5'-flanking regions of cytochrome p450 (CYP)3A4 and CYP2C9 genes

PT respectively -

XX

XX Disclosure; Page 3; 47pp; English.

XX

XX The present invention relates to PCR primer pairs for amplifying

CC and sequence determination oligonucleotides for detecting single

CC nucleotide polymorphisms (SNPs) in the 5'-flanking regions of human

CC cytochrome p450 (CYP) genes encoding CYP3A4 or CYP2C9. The SNPs

CC correspond to position 461 of a defined 1345 base pair sequence for

CC CYP3A4 or position 957, 1049, 1164, 1526, 1661 and 1662 of a 2438 base

CC pair sequence for CYP2C9. The PCR primers are useful for amplifying

CC the CYP sequences and the oligonucleotides are useful for detecting

CC SNPs in the 5'-flanking regions of the CYP3A4 or CYP2C9 genes.

CC ABK68747-ABK68750 represent previously published oligonucleotides

CC for detecting a polymorphism in the CYP3A4 gene.

XX

XX Sequence 19 BP; 7 A; 3 C; 9 G; 0 U; 0 other;

SQ Query Match 85.6%; Score 15.4; DB 24; Length 19;

Best Local Similarity 94.1%; Pred. No. 5.3e+02;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 ACAAGGCGAGCAGACAG 18

|||||

Db 1 ACAAGGCGAGCAGACAG 17

RESULT 14

ABQ71128

ID ABQ71128 standard; DNA; 677 BP.

AC ABQ71128;

XX

XX 29-AUG-2002 (first entry)

DT

XX

XX Listeria monocytogenes 4b specific contig55.

DE

XX

XX Antibacterial; Listeria; food contamination; mutational analysis;

KW infection; ds.

XX

XX Listeria monocytogenes 4b.

OS

XX

XX WO200228891-A2.

PN

XX

XX 11-APR-2002.

PD

XX

XX 04-OCT-2001; 2001WO-FR03061.

PF

XX

XX 04-OCT-2000; 2000FR-0012697.

PR

XX

XX (INSP ) INST PASTEUR.

PA

XX

XX (CNRS ) CNRS CENT NAT RECH SCI.

PA

XX

XX Kunst F, Glaser P;

PI

XX

XX WPI; 2002-332479/37.

DR

XX

XX New genomic sequences from Listeria species, useful for detection,

PT treatment and prevention of infection, also related polypeptides,

PT antibodies and modulators -

PT

XX

XX Claim 22; SEQ ID 3941; 180pp; French.

PS

XX

XX The present invention relates to nucleic acid sequences

CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes

CC and primers for identification and/or detection of Listeria (e.g. as

CC contaminants in foods, or mutational analysis) and for analysis of

CC gene expression. Proteins encoded by the nucleic acid sequences can be

CC used to screen for compounds that modulate gene expression, replication

CC and pathogenicity of Listeria (potential therapeutic agents), also for

CC treating infections by Listeria, and are useful as immunogens in

CC anti-Listeria vaccines.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic format

CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX

XX Sequence 677 BP; 145 A; 133 C; 206 G; 193 T; 0 other;

SQ Query Match 85.6%; Score 15.4; DB 24; Length 677;

Best Local Similarity 94.1%; Pred. No. 6.2e+02;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 ACAAGGCGAGCAGACAG 18

|||||

Db 60 AGAGGGCGAGCAGACAG 76

RESULT 15

AXX23520

ID AXX23520 standard; DNA; 45546 BP.

XX

XX AAX23520;

AC

XX 23-JUN-1999 (first entry)

DT

XX

XX Human kidney aminopeptidase P genomic DNA fragment 4.

DE

XX

XX Aminopeptidase; human; AMP; gene therapy; treatment; AMP-deficiency;

KW prenatal diagnosis; angioedema; antihypertensive agent; atherosclerosis;

KW arterial stenosis; industrial protein feed; malabsorption syndrome;

KW proteinaceous waste degradation; additive; immunohistochemistry; ss.

XX

XX Homo sapiens.

OS

XX

XX WO9911799-A2.

PN

XX

XX 11-MAR-1999.

PD

XX

XX 02-SEP-1998; 98WO-US18426.

PF

XX

XX 02-SEP-1997; 97US-0057854.

PR

XX

XX (MEDI-) MEDICAL COLLEGE GEORGIA RES INST.

PA

XX

XX Ryan JW, Sprinkle TJC, Venema RC;

PI

XX

XX WPI; 1999-205193/17.

DR

XX

XX Nucleic acid encoding human aminopeptidase P

PT

XX

XX Claim 13; Page 165-192; 201pp; English.

PS

XX

XX This invention describes the isolation of a novel human aminopeptidase P

CC (AMP). This protein is used to produce recombinant AMP and can be used

CC for gene therapy for treating AMP-deficiency conditions. Its fragments

CC are used as primers and probes to identify patients with homozygous and

CC heterozygous AMP deficiency, including prenatal diagnosis (patients

CC defective in AMP are at risk of developing angioedema if treated with

CC angiotensin-converting enzyme inhibitors), also as antisense inhibitors

CC in cases of excessive AMP expression. The product of the invention is

CC also used to identify AMP-expressing sequences in other animals and to

CC generate transgenic animals, and comparisons of genomic sequences are

CC used to detect mutations. AMP inhibitors are potentially useful as

CC antihypertensive agents and to prevent or treat arterial (re)stenosis

CC or atherosclerosis. The structure of AMP is used to design synthetic

CC substrates, e.g. for use in AMP assays. AMP, which hydrolyzes N-terminal

CC imido bonds, can be used to degrade industrial protein feeds to free

CC amino acids, to degrade proteinaceous wastes, as additives in enzyme

CC formulations used to treat malabsorption syndrome and for studying its

CC biological role. Antibodies against AMP are used in immunohistochemical

CC methods to study AMP distribution.

XX

SQ Sequence 45546 BP; 12027 A; 11359 C; 11380 G; 10780 T; 0 other;

Query Match 85.6%; Score 15.4; DB 20; Length 45546;

Best Local Similarity 94.1%; Pred. No. 7.6e+02;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACAGGGCAGGACAGA 17

| | | | | | | | | | | | | | | |

Db 565 GGCAGGGCAGGACAGA 581

Search completed: August 27, 2003, 18:15:44

Job time : 142.192 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 27, 2003, 18:02:39 ; Search time 69.9615 Seconds  
(without alignments)  
214.504 Million cell updates/sec

Title: US-10-085-612-2

Perfect score: 0  
Sequence: 1 cgattcttgctgctgcagctgcagccoccy 34

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database : Issued Patents.NA.\*  
1: /cgn2\_6/prodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/prodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/prodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/prodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/prodata/1/ina/PCTUS.COMB.seq.\*  
6: /cgn2\_6/prodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22.8	67.1	1345	3 US-09-372-339-1	Sequence 1, Appli
2	22.8	67.1	1345	3 US-09-372-339-2	Sequence 2, Appli
3	22.8	67.1	1345	4 US-09-144-367-3	Sequence 3, Appli
C 4	20	58.8	357	2 US-08-294-143-1	Sequence 1, Appli
C 5	20	58.8	357	3 US-09-256-331-1	Sequence 1, Appli
C 6	20	58.8	357	4 US-09-593-483-1	Sequence 1, Appli
C 7	20	58.8	829	2 US-08-294-143-3	Sequence 3, Appli
C 8	20	58.8	829	3 US-09-256-331-3	Sequence 3, Appli
C 9	20	58.8	829	4 US-09-593-483-3	Sequence 3, Appli
C 10	19.2	56.5	34185	4 US-09-545-481-3	Sequence 3, Appli
C 11	19	55.9	286	4 US-09-313-294A-2795	Sequence 2795, Ap
C 12	19	55.9	4550	3 US-09-103-663-35	Sequence 35, Appl
C 13	19	55.9	4960	4 US-09-907-843-3	Sequence 3, Appli
C 14	18.8	55.3	238	1 US-07-903-466-32	Sequence 32, Appl
C 15	18.8	55.3	238	5 PCT-US93-05794-32	Sequence 32, Appl
C 16	18.8	55.3	317	3 US-08-964-268-20	Sequence 20, Appl
C 17	18.8	55.3	317	4 US-09-105-234-20	Sequence 20, Appl
C 18	18.8	55.3	1767	1 US-07-903-466-2	Sequence 2, Appli
C 19	18.8	55.3	1767	5 PCT-US93-05794-2	Sequence 19, Appl
C 20	18.8	55.3	2487	4 US-09-257-894-19	Sequence 19, Appl
C 21	18.8	55.3	2772	4 US-09-257-894-12	Sequence 12, Appl
C 22	18.8	55.3	3018	1 US-07-903-466-1	Sequence 1, Appli
C 23	18.8	55.3	3018	5 PCT-US93-05794-1	Sequence 1, Appli
C 24	18.8	55.3	15664	1 US-08-402-282-3	Sequence 3, Appli
C 25	18.8	55.3	15664	1 US-08-508-004-3	Sequence 3, Appli
C 26	18.8	55.3	15664	1 US-08-402-066-3	Sequence 3, Appli
C 27	18.8	55.3	15664	1 US-08-402-068-3	Sequence 3, Appli

```

28 18.4 54.1 435 4 US-09-397-787-162 Sequence 162, App
29 18.4 54.1 705 4 US-09-328-352-2637 Sequence 2637, Ap
30 18.4 54.1 1335 4 US-09-450-790A-18 Sequence 18, Appli
31 18.4 54.1 1406 1 US-08-745-269-1 Sequence 1, Appli
32 18.4 54.1 1406 2 US-08-157-185-1 Sequence 1, Appli
33 18.4 54.1 1406 3 US-08-281-526B-1 Sequence 1, Appli
34 18.4 54.1 1406 4 US-09-450-797-1 Sequence 1, Appli
35 18.4 54.1 1406 4 US-09-450-790A-1 Sequence 1, Appli
36 18.4 54.1 1406 4 US-09-332-837-1 Sequence 1362, Ap
37 18.4 54.1 1406 4 US-09-016-434-1362 Sequence 1, Appli
38 18.4 54.1 1406 5 PCT-US93-10553-1 Sequence 8, Appli
39 18.4 54.1 1417 2 US-08-428-243-8 Sequence 8, Appli
40 18.4 54.1 1417 5 PCT-US93-10301-8 Sequence 5, Appli
41 18.4 54.1 2031 4 US-09-693-147-5 Sequence 41, Appli
42 18.4 54.1 5640 4 US-09-620-312D-41 Sequence 41, Appli
43 18.2 53.5 723 4 US-09-252-991A-10004 Sequence 10004, A
44 18.2 53.5 825 4 US-09-252-991A-14839 Sequence 14839, A
45 18.2 53.5 1050 4 US-09-252-991A-14525 Sequence 14525, A

```

#### ALIGNMENTS

##### RESULT 1

```

US-09-372-339-1
; Sequence 1, Application US/09372339
; Patent No. 6174684
; GENERAL INFORMATION:
; APPLICANT: Rebbeck, Timothy
; APPLICANT: Felix, Carolyn
; TITLE OF INVENTION: CYP3A4 NFSE Variant and Methods of Use Therefor
; FILE REFERENCE: PENN-0695
; CURRENT APPLICATION NUMBER: US/09/372,339
; CURRENT FILING DATE: 1999-08-11
; EARLIER FILING DATE: 1998-08-14
; EARLIER FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1345
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-372-339-1

```

```

Query Match Similarity 57.18; Score 22.8; DB 3; Length 1345;
Best Local Similarity 92.3%; Freq. No. 4.2;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 9 TGGTACTGGCTGCAGCTGCAGCCCG 34
      |||||
Db 1035 TGGTACTGGCTGCAGCTGCAGCCCTG 1060

```

##### RESULT 2

```

US-09-372-339-2
; Sequence 2, Application US/09372339
; Patent No. 6174684
; GENERAL INFORMATION:
; APPLICANT: Rebbeck, Timothy
; APPLICANT: Felix, Carolyn
; TITLE OF INVENTION: CYP3A4 NFSE Variant and Methods of Use Therefor
; FILE REFERENCE: PENN-0695
; CURRENT APPLICATION NUMBER: US/09/372,339
; CURRENT FILING DATE: 1999-08-11
; EARLIER FILING DATE: 1998-08-14
; EARLIER FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1345
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-372-339-2

```

Query Match 67.1%; Score 22.8; DB 3; Length 1345;  
Best Local Similarity 92.3%; Pred. No. 4.2;  
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 9 TGCTACTGGCTGCAGCTGCAGCCCG 34  
|||||  
DB 1035 TGCTACTGGCTGCAGCTGCAGCCCG 1060

RESULT 3  
US-09-144-367-3  
; Sequence 3, Application US/09144367  
; Patent No. 6432639  
; GENERAL INFORMATION:  
; APPLICANT: Lichter, Jay  
; APPLICANT: Guido, Marco  
; TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4  
; FILE REFERENCE: SEQ-12P  
; CURRENT APPLICATION NUMBER: US/09/144,367  
; CURRENT FILING DATE: 1998-08-31  
; PRIOR APPLICATION NUMBER: 60/058,612  
; PRIOR FILING DATE: 1997-09-10  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 1345  
; TYPE: DNA  
; ORGANISM: H. sapiens  
; FEATURE:  
; NAME/KEY: Other  
; LOCATION: (0)....(0)  
US-09-144-367-3

Query Match 67.1%; Score 22.8; DB 4; Length 1345;  
Best Local Similarity 92.3%; Pred. No. 4.2;  
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 9 TGCTACTGGCTGCAGCTGCAGCCCG 34  
|||||  
DB 1035 TGCTACTGGCTGCAGCTGCAGCCCG 1060

RESULT 4  
US-08-294-143-1/c  
; Sequence 1, Application US/08294143  
; Patent No. 5874231  
; GENERAL INFORMATION:  
; APPLICANT: NAHUM SONENBERG  
; APPLICANT: ARNIM PAUSE  
; APPLICANT: JOE B. HARFORD  
; APPLICANT: VINCENT J. MILES  
; TITLE OF INVENTION: METHODS FOR TREATING  
; TITLE OF INVENTION: HORMONE DISORDERS  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; CITY: Suite 4700  
; STATE: Los Angeles  
; COUNTRY: California  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/294,143  
; FILING DATE: August 22, 1994  
; CLASSIFICATION: 424

; PRIOR APPLICATION DATA: including application  
; PRIOR APPLICATION DATA: Described below:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 202/060  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 357 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-294-143-1

Query Match 58.8%; Score 20; DB 2; Length 357;  
Best Local Similarity 82.1%; Pred. No. 41;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 6 CTTTGCTACTGGCTGCAGCTGCAGCCCG 33  
|||||  
DB 35 CTTGGGCTGCTGCAGCTGCAGCCCG 8

RESULT 5  
US-09-256-331-1/c  
; Sequence 1, Application US/09256331  
; Patent No. 611077  
; GENERAL INFORMATION:  
; APPLICANT: NAHUM SONENBERG  
; APPLICANT: ARNIM PAUSE  
; APPLICANT: JOE B. HARFORD  
; APPLICANT: VINCENT J. MILES  
; TITLE OF INVENTION: METHODS FOR TREATING  
; TITLE OF INVENTION: HORMONE DISORDERS  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; CITY: Suite 4700  
; STATE: Los Angeles  
; COUNTRY: California  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/256,331  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/294,143  
; FILING DATE: August 22, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 202/060  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 357 base pairs  
; TYPE: nucleic acid

; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-593-483-1

Query Match 58.8%; Score 20; DB 3; Length 357;  
Best Local Similarity 82.1%; Pred. No. 41;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 6 CTTTGCTACTGGCTGCAGCTGCAGCCCC 33  
||| | ||||| ||||| |||||  
Db 35 CTTGGGCTCTGGCTGCAGCTGCAGCCCC 8

## RESULT 6

US-09-593-483-1/c  
; Sequence 1, Application US/09593483  
; Patent No. 6410715  
; GENERAL INFORMATION:  
; APPLICANT: NAHUM SONENBERG  
; ARNIM PAUSE  
; JOE B. HARFORD  
; VINCENT J. MILES

TITLE OF INVENTION: METHODS FOR TREATING  
HORMONE DISORDERS

NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/593,483  
FILING DATE: 14-Jun-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/294,143  
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327

REFERENCE/DOCKET NUMBER: 202/060

TELECOMMUNICATION INFORMATION:

TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 357 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

Query Match 58.8%; Score 20; DB 4; Length 357;  
Best Local Similarity 82.1%; Pred. No. 41;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 6 CTTTGCTACTGGCTGCAGCTGCAGCCCC 33  
||| | ||||| ||||| |||||  
Db 35 CTTGGGCTCTGGCTGCAGCTGCAGCCCC 8

## RESULT 7

US-08-294-143-3/c  
; Sequence 3, Application US/08294143  
; Patent No. 5874231  
; GENERAL INFORMATION:  
; APPLICANT: NAHUM SONENBERG  
; APPLICANT: ARNIM PAUSE  
; APPLICANT: JOE B. HARFORD  
; APPLICANT: VINCENT J. MILES

TITLE OF INVENTION: METHODS FOR TREATING  
HORMONE DISORDERS

NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/294,143  
FILING DATE: August 22, 1994

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below:

ATTORNEY/AGENT INFORMATION:

NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327

REFERENCE/DOCKET NUMBER: 202/060

TELECOMMUNICATION INFORMATION:

TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 829 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

TOPOLOGY: linear

US-08-294-143-3

Query Match 58.8%; Score 20; DB 2; Length 829;  
Best Local Similarity 82.1%; Pred. No. 48;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 6 CTTTGCTACTGGCTGCAGCTGCAGCCCC 33  
||| | ||||| ||||| |||||  
Db 62 CTTGGGCTCTGGCTGCAGCTGCAGCCCC 35

## RESULT 8

US-09-256-331-3/c  
; Sequence 3, Application US/09256331  
; Patent No. 6111077  
; GENERAL INFORMATION:  
; APPLICANT: NAHUM SONENBERG  
; APPLICANT: ARNIM PAUSE  
; APPLICANT: JOE B. HARFORD  
; APPLICANT: VINCENT J. MILES

TITLE OF INVENTION: METHODS FOR TREATING  
HORMONE DISORDERS

NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
Suite 4700





LENGTH: 286  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: misc\_feature  
OTHER INFORMATION: Incyte ID No. 6476212 700553275H1  
US-09-313-294A-2795

Query Match 55.9%; Score 19; DB 4; Length 286;  
Best Local Similarity 81.5%; Pred. No. 97;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 ATTCTTGTCTACTGCTGCAGCTGCAG 29  
||||||| | |||||||||  
DB 17 ATTCTTTGGCAGCTGCTGCAGCTGCAG 43

RESULT 12  
US-09-103-663-35/c  
Sequence 35, Application US/09103663D  
Patent No. 6171803  
GENERAL INFORMATION:  
APPLICANT: Kinnet et al.  
TITLE OF INVENTION: Isolation, characterization, and use of the human beta  
TITLE OF INVENTION: subunit of the high affinity receptor for  
TITLE OF INVENTION: immunoglobulin E.  
FILE REFERENCE: 50490  
CURRENT APPLICATION NUMBER: US/09/103,663D  
CURRENT FILING DATE: 1998-06-23  
EARLIER APPLICATION NUMBER: 07/869,933  
EARLIER FILING DATE: 1992-04-16  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 35

LENGTH: 4550  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (578)  
OTHER INFORMATION: n represents a, c, t or g.  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (735)  
OTHER INFORMATION: n represents a, c, t or g.  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1362)  
OTHER INFORMATION: n represents a, c, t or g  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (2479)  
OTHER INFORMATION: n represents a, c, t or g.  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (2517)  
OTHER INFORMATION: n represents a, c, t or g.  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (2536)  
OTHER INFORMATION: n represents a, c, t or g.  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (2549)..(2552)  
OTHER INFORMATION: n represents a, c, t or g.  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (2612)..(2614)  
OTHER INFORMATION: n represents a, c, t or g.  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (2633)  
OTHER INFORMATION: n represents a, c, t or g.

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (2920)  
OTHER INFORMATION: n represents a, c, t or g.  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (3069)  
OTHER INFORMATION: n represents a, c, t or g.  
US-09-103-663-35

Query Match 55.9%; Score 19; DB 3; Length 4550;  
Best Local Similarity 81.5%; Pred. No. 1.5e+02;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 ATTCTTGTCTACTGCTGCAGCTGCAG 29  
||||||| | |||||||||  
DB 1872 ATTCTTAGATAGTGGCTGGAGCTGTAG 1846

RESULT 13  
US-09-907-843-3  
Sequence 3, Application US/09907843  
Patent No. 6440739  
GENERAL INFORMATION:  
APPLICANT: C. Frank Bennett  
APPLICANT: Susan M. Freier  
TITLE OF INVENTION: ANTISENSE MODULATION OF GLIOMA-ASSOCIATED ONCOGENE-2 EXPRESSIO  
FILE REFERENCE: RTS-0279  
CURRENT APPLICATION NUMBER: US/09/907,843  
CURRENT FILING DATE: 2001-07-17  
NUMBER OF SEQ ID NOS: 87  
SEQ ID NO 3  
LENGTH: 4960  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (572)...(4348)  
US-09-907-843-3

Query Match 55.9%; Score 19; DB 4; Length 4960;  
Best Local Similarity 81.5%; Pred. No. 1.6e+02;  
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 TCTTTGCTACTGCTGCAGCTGCAGC 30  
||||||| | |||||||||  
DB 383 TCCTTTGGTGGCTGCAGCGGCAGC 409

RESULT 14  
US-07-903-466-32/c  
Sequence 32, Application US/07903466  
Patent No. 5395767  
GENERAL INFORMATION:  
APPLICANT: Murnane, John P.  
APPLICANT: Painter, Robert B.  
APPLICANT: Kapp, Leon N.  
APPLICANT: Yu, Loh C.  
TITLE OF INVENTION: Gene for Ataxia-Telangiectasia  
TITLE OF INVENTION: Complimentation Group D (ATDC)  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: Steuart Street Tower, 18th Fl., One Market  
STREET: Plaza  
CITY: San Francisco  
STATE: California  
COUNTRY: San Francisco  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/903,466  
FILING DATE: 19920622  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: 91-077-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-777-9275  
TELEFAX: 415-543-4219  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 238 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-07-903-466-32

Query Match 55.3%; Score 18.8; DB 1; Length 238;  
Best Local Similarity 76.7%; Pred. No. 1.1e-02;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CGATTCTTTGCTACTGGCTGCAGCTGCAGC 30  
||| | ||||| ||||| ||| |||||  
Db 145 CGACCCGTTGCTCTCTGGAGGCATCTGCAGC 116

## RESULT 15

PCT-US93-05794-32/c  
Sequence 32, Application PC/TUS9305794  
GENERAL INFORMATION:  
APPLICANT: The Regents of the University of California  
TITLE OF INVENTION: Gene for Ataxia-Telangiectasia  
TITLE OF INVENTION: Complementatation Group D (ATDC)  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 177 Post Street, Suite 800  
CITY: San Francisco  
STATE: California  
COUNTRY: San Francisco  
ZIP: 94108-4731  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/05794  
FILING DATE: 19930618  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/903,466  
FILING DATE: 22-JUN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: 91-077-1 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-421-4973  
TELEFAX: 415-421-1663  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 238 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO

ANTI-SENSE: NO  
PCT-US93-05794-32

Query Match 55.3%; Score 18.8; DB 5; Length 238;  
Best Local Similarity 76.7%; Pred. No. 1.1e+02;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CGATTCTTTGCTACTGGCTGCAGCTGCAGC 30  
||| | ||||| ||||| ||| |||||  
Db 145 CGACCCGTTGCTCTCTGGAGGCATCTGCAGC 116

Search completed: August 27, 2003, 19:42:02  
Job time : 70.9615 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 27, 2003, 18:01:39 ; Search time 2272.12 Seconds  
(without alignments)  
363.693 Million cell updates/sec

Title: us-10-085-612-2

Perfect score: 34

Sequence: 1 cgattcttgtaactgctgcagctgcagccccc 34

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estha.\*

2: em\_esthum.\*

3: em\_estnu.\*

4: em\_estov.\*

5: em\_estpl.\*

6: em\_estro.\*

7: em\_estro.\*

8: em\_estro.\*

9: gb\_estti.\*

10: gb\_estti.\*

11: gb\_estti.\*

12: gb\_estti.\*

13: gb\_estti.\*

14: gb\_estti.\*

15: em\_estfun.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_pln.\*

20: em\_gss\_vrt.\*

21: em\_gss\_fun.\*

22: em\_gss\_mam.\*

23: em\_gss\_mus.\*

24: em\_gss\_pro.\*

25: em\_gss\_rod.\*

26: em\_gss\_phg.\*

27: em\_gss\_vrl.\*

28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	24.4	71.8	802	29	BZ571326
c 2	23.4	68.8	910	29	AL211728 Tetraodon
c 3	23	67.6	438	12	BM037193
4	23	67.6	660	9	AA739600

c	5	22.6	66.5	704	29	CC011462
	6	22.2	65.3	603	9	AA803400
	7	22.2	65.3	611	9	AA538677
	8	22	64.7	180	28	BH613513
	9	22	64.7	196	28	BH214265
	10	21.8	64.1	320	14	T60367
	11	21.8	64.1	480	9	AJ478560
	12	21.8	64.1	509	9	AJ432574
	13	21.8	64.1	667	10	BE730257
	14	21.8	64.1	687	12	BJ219877
c	15	21.8	64.1	904	29	CNS04FD7
	16	21.8	64.1	946	10	BG343303
	17	21.8	64.1	972	13	BU859835
	18	21.8	64.1	991	12	BM472533
	19	21.8	64.1	1043	12	BM548921
c	20	21.8	64.1	1201	29	CNS05KQ1
	21	21.6	63.5	549	12	BM003553
c	22	21.6	63.5	681	12	BM621560
	23	21.6	63.5	688	12	BM578815
c	24	21.6	63.5	720	28	BM575468
	25	21.4	62.9	368	28	BM253181
	26	21.4	62.9	422	9	AW501182
	27	21.4	62.9	485	13	BM282025
c	28	21.4	62.9	518	9	AI568540
	29	21.4	62.9	527	12	BI868034
	30	21.4	62.9	618	10	BF431634
	31	21.4	62.9	623	12	BI866970
	32	21.4	62.9	649	10	BG424809
	33	21.4	62.9	663	10	PG335538
	34	21.4	62.9	683	12	BI562582
	35	21.4	62.9	687	12	BG911254
	36	21.4	62.9	713	12	BI910096
	37	21.4	62.9	718	12	BI545634
	38	21.4	62.9	723	10	BG717956
	39	21.4	62.9	777	10	BF984839
	40	21.4	62.9	798	10	BG717465
	41	21.4	62.9	848	13	BU594473
	42	21.4	62.9	903	13	BU500183
	43	21.4	62.9	915	14	CD358712
c	44	21.4	62.9	929	29	CNS04LLZ
	45	21.4	62.9	943	13	BU193557

## ALIGNMENTS

RESULT 1  
BZ571326/c  
LOCUS msh2\_1835.xl msh Pseudomonas aeruginosa genomic clone msh2\_1835,  
DEFINITION genomic survey sequence.  
ACCESSION BZ571326  
VERSION BZ571326.1 GI:27206387  
KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
REFERENCE 1 (bases 1 to 802)  
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
Burns, J.L., Kaul, R. and Olsen, M.V.  
TITLE Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
JOURNAL J. Bacteriol., (2002), In press  
COMMENT Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: shotgun.  
Location/Qualifiers

```

source
1. .802
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"
/clone="msb2_1835"
/clone_lib="msh"
/notes="Environmental isolate. Whole genomic shotgun library."
BASE COUNT      139 a      218 c      164 g      228 t      53 others
ORIGIN

Query Match      71.8%; Score 24.4; DB 29; Length 802;
Best Local Similarity 82.4%; Pred. No. 1.6e+02;
Matches 28; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CGATCTTTGCTACTGGCTGCGAGCTGCAGCCGCG 34
||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 348 CGATCTTTGCTACTGGCTGCGAGCTGCAGCCGCG 315

RESULT 2
CNS02SD3/c
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
161322 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION
AL211728.1 GI:7870547
VERSION
AL211728.1
KEYWORDS
GSS: genome survey sequence.
SOURCE
Tetraodon nigroviridis
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontoidea; Tetraodontidae; Tetraodon.
REFERENCE
1
AUTHORS
Roest Crolius H., Jaillon O., Dasilva C., Bouneau L., Fisher C.,
Bernot A., Fizames C., Wincker P., Brottier P., Quetier F.,
Saurin W. and Weissenbach J.
TITLE
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
JOURNAL
Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE
20296633
PUBMED
10835845
REFERENCE
2
AUTHORS
Roest Crolius H., Jaillon O., Dasilva C., Ozouf-Costaz C.,
Fizames C., Fischer C., Bouneau L., Billault A., Quetier F.,
Saurin W., Bernot A. and Weissenbach J.
TITLE
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL
Genome Res. 10 (7), 939-949 (2000)
MEDLINE
20359837
PUBMED
10899143
REFERENCE
3 (bases 1 to 910)
GENSCOPE
Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetraodon.
FEATURES
location/Qualifiers
1. .910
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="161322"
/clone_lib="G"
/notes="Genoscope sequence ID : COAG161DE11SP1-end :
PUC-Ori"
source

BASE COUNT      138 a      78 c      100 g      122 t
ORIGIN

Query Match      67.6%; Score 23; DB 12; Length 438;

FEATURES
source
1. .438
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:5377536"
/sex="female"
/dev_stage="4-5 month"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Gong zebrafish ovary"
/notes="Organ: ovary (pooled); vector: pBluescript SK-;
Site1: XhoI; Site2: EcoRI; Poly A+ RNA was isolated from
the ovaries of 2 female adult zebrafish (4-5 month old).
cDNAs were made using oligo-dT primers and inserted into
lambda ZAP II vector (Stratagene) by Dr. Z. Gong, in vivo
mass-excisued to pBluescript SK- following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Zhiyuan Gong for further information on
this library (National University of Singapore,
Department of Biological Sciences, Lower Kent Ridge Road,
Singapore 119260)."
BASE COUNT      138 a      78 c      100 g      122 t
ORIGIN

Query Match      67.6%; Score 23; DB 12; Length 438;

```



GSS.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Arabidopsis thaliana (thale cress)  
 Arabidopsis thaliana  
 Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:

REFERENCE  
AUTHORS : eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 196)  
Alonso,J.M., Leisner,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab  
,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.  
, Zimmerman,J. and Ecker,J.R.

TITLE A Sequence-Indexed Library of Insertion Mutations in the

JOURNAL Arabidopsis Genome

COMMENT Unpublished

Contact: Joseph R. Ecker  
Salk Institute Genomic Analysis Laboratory (SIGnAL)  
The Salk Institute for Biological Studies  
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
Tel: 858 453 4100 x1752  
Fax: 858 558 6379

Email: ecker@salk.edu

This is single pass sequence recovered from the left border of  
TDNA. This sequence lies within an intron of At1g20480.

Class: TDNA tagged.

FEATURES Location/Qualifiers

source

1. .196  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia 0"  
/db\_xref="taxon:3702"  
/clone="SALK\_010491"  
/clone\_lib="Arabidopsis thaliana TDNA insertion lines"  
/note="PCR was performed on Arabidopsis thaliana lines  
each of which contains one or more TDNA insertion  
elements. The resultant fragment for each line was  
directly sequenced to determine the genomic sequence at  
the site of insertion. Details of the protocols used can  
be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)"

BASE COUNT 43 a 48 c 47 g 58 t

ORIGIN

Query Match 64.7%; Score 22; DB 28; Length 196;

Best Local Similarity 83.3%; Pred. No. 8e+02; 5; Indels 0; Gaps 0;  
Matches 25; Conservative 0; Mismatches 5;

Qy 5 TCTTTGCTACTGCTGACGCTGCAGCCCG 34

Db 131 TTTTGTCTATTGCTGACGCTGCAGCCCG 160

RESULT 10

T60367  
LOCUS yb90h09.r1 Striatogene liver (#937224) Homo sapiens cDNA clone  
DEFINITION IMAGE:78437 5' similar to gb:J04449 CYTOCHROME P450 IIIA4 (HUMAN);,  
mRNA sequence.

ACCESSION T60367

VERSION T60367.1 GI:663404

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 320)

1 Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,  
Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins  
,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore  
,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,  
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E.,  
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)

97044478

PUBMED 889549

COMMENT Contact: Wilson RK

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: estewatson.wustl.edu  
High quality sequence stops: 209

Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: M13RP1

High quality sequence stop: 209.

FEATURES

source

1. .320  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="GDB:498242"  
/db\_xref="taxon:9606"  
/clone="IMAGE:78497"  
/sex="male"  
/dev\_stage="49 years old"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/clone\_lib="Striatogene liver (#937224)"  
/note="Organ: liver; Vector: Bluescript SK; Site:1: EcoRI  
; Site:2: XhoI; Cloned unidirectionally. Primer: Oligo  
dT. Hepatotomy from normal male caucasian. Average insert  
size: 1.1 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5'  
GAATTCGCGACGAG 3' -3' adaptor sequence: 5'  
CTCGAGTTTCTTTTCTTTT 3'"

BASE COUNT 87 a 82 c 77 g 66 t 8 others

ORIGIN

Query Match 64.1%; Score 21.8; DB 14; Length 320;

Best Local Similarity 88.5%; Pred. No. 1.1e+03;  
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 9 TGTCTGCTGCTGACGCTGCAGCCCG 34

Db 47 TGCTACTGCTGCANCTCCAGCCCTG 72

RESULT 11

AJ478560

LOCUS

DEFINITION AJ478560 S00011 Hordeum vulgare cDNA clone S0001100114G04F1, mRNA  
sequence.

ACCESSION AJ478560

VERSION AJ478560.1 GI:21194515

KEYWORDS EST.

SOURCE Hordeum vulgare

ORGANISM Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae  
; Triticeae; Hordeum.  
1 (bases 1 to 480)

REFERENCE

AUTHORS Saren,A.-M., Tanskanen,J., Paulin,L. and Schulman,A.H.

TITLE Barley EST's

JOURNAL Unpublished

COMMENT Contact: Schulman AH

Institute of Biotechnology

University of Helsinki

P.O.Box 56 (Viikinkaari 6A), University of Helsinki FIN-00014,

Finland.

FEATURES

source

1. .480  
Location/Qualifiers  
/organism="Hordeum vulgare"  
/mol\_type="mRNA"  
/db\_xref="taxon:4513"  
/clone="S0001100114G04F1"  
/dev\_stage="Developing seed"  
/clone\_lib="S00011"  
/note="12,15,18 days after pollination"

BASE COUNT 94 a 171 c 137 g 78 t

ORIGIN

Query Match 64.1%; Score 21.8; DB 9; Length 480;

Best Local Similarity 78.8%; Pred. No. 1.2e+03;  
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;



```

Qy 2 GATTCTTTGCTAGCTGCTGAGTGCAGCCCG 34
    ||| | | ||| ||||| ||||| ||
Db 311 GATCGGCTCTACAGCTGCTGAGTGCAGCG 343

RESULT 12
AJ432574
LOCUS AJ432574 509 bp mRNA linear EST 15-MAR-2002
DEFINITION AJ432574 S00011 Hordeum vulgare cDNA clone S0001100156E11F1, mRNA
sequence.
ACCESSION AJ432574
VERSION AJ432574.1 GI:19521026
KEYWORDS EST.
SOURCE Hordeum vulgare
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Hordeum.
REFERENCE Saren,A.-M., Tanskanen,J., Paulin,L. and Schulman,A.H.
AUTHORS Barley EST's
TITLE Unpublished
JOURNAL
COMMENT Contact: Schulman AH
Institute of Biotechnology
University of Helsinki
P.O.Box 56 (Viikinkaari 6A), University of Helsinki FIN-00014,
Finland.

FEATURES
source
Query Match 64.1%; Score 21.8; DB 9; Length 509;
Best Local Similarity 78.8%; Pred. No. 1.2e+03;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 GATTCTTTGCTAGCTGCTGAGTGCAGCCCG 34
    ||| | | ||| ||||| ||||| ||
Db 303 GATCGGCTCTACAGCTGCTGAGTGCAGCG 335

RESULT 13
BE730257
LOCUS BE730257 667 bp mRNA linear EST 15-SEP-2000
DEFINITION 601563625F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3832999 5',
mRNA sequence.
ACCESSION BE730257
VERSION BE730257.1 GI:10144249
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgraphs-remail.nih.gov
Tissue Procurement: ATCC/DCID/DTF
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: L1CM511 row: b column: 08
High quality sequence stop: 667.
Location/Qualifiers
1. .667
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3832999"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_20"
/notes="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

[illegible]

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: August 27, 2003, 18:00:44 ; Search time 536.538 Seconds  
(without alignments)  
1372.452 Million cell updates/sec  
Title: US-10-085-612-1  
Perfect score: 18  
Sequence: 1 gacaaggcaggacagag 18  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2888711 seqs, 20454813386 residues  
Total number of hits satisfying chosen parameters: 5777422  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.man.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB ID	Description
1	18	100.0	144869	2	AC114884 Rattus no
2	18	100.0	178857	2	AC114881 Rattus no
3	18	100.0	188490	2	AC118733 Mus muscu
4	18	100.0	196662	2	AC129737 Rattus no
5	18	100.0	209706	2	AC101931 Mus muscu
6	18	100.0	217691	2	AC103052 Rattus no
7	18	100.0	244620	2	AC119882 Mus muscu
8	17	94.4	47589	2	AC124053 Mus muscu
9	17	94.4	116608	2	AC121356 Mus muscu
10	17	94.4	151613	2	AC144886 Bos tauru
11	17	94.4	163475	10	AC026761 Mus muscu
12	17	94.4	173362	2	AC145188 Bos tauru
13	17	94.4	196351	2	AC138402 Mus muscu
14	17	94.4	205621	10	AC079082 Mus muscu
15	17	94.4	217181	2	AC132033 Rattus no
16	17	94.4	219937	2	AC135670 Mus muscu
17	17	94.4	237330	2	AC098104 Rattus no
18	17	94.4	238658	10	AC079818 Mus muscu
19	17	94.4	239553	2	AC123201 Rattus no
20	17	94.4	271890	2	AC106243 Rattus no
21	17	94.4	287536	2	AC093993 Rattus no
22	17	94.4	347924	2	AC131229 Mus muscu
23	16.4	91.1	29	6	AR222903 Sequence
24	16.4	91.1	532	4	CHI292058
25	16.4	91.1	573	9	HS295982 Homo sapi
26	16.4	91.1	590	11	G95069
27	16.4	91.1	1345	6	AR142140
28	16.4	91.1	8088	4	CHBIG
29	16.4	91.1	32735	9	AC006124 Homo sapi
30	16.4	91.1	38016	9	AC138433 Homo sapi
31	16.4	91.1	38235	9	AC004559 Homo sapi
32	16.4	91.1	60271	2	AC110286 Homo sapi
33	16.4	91.1	62915	2	AL929091_16 Continuation (17 o
34	16.4	91.1	64706	2	AC136358 Homo sapi
35	16.4	91.1	75236	6	AX600586 Sequence
36	16.4	91.1	75761	9	AL359733 Human DNA
37	16.4	91.1	88592	10	AL732630 Mouse DNA
38	16.4	91.1	103706	9	AL354891 Human DNA
39	16.4	91.1	104871	9	AC004584 Homo sapi
40	16.4	91.1	106041	10	AC074359 Mus muscu
41	16.4	91.1	108195	9	AC074129 Homo sapi
42	16.4	91.1	110000	2	AC108631 Rattus no
43	16.4	91.1	110000	2	Continuation (13 o
44	16.4	91.1	118172	9	AC015912 Homo sapi
45	16.4	91.1	123778	9	AC069294 Homo sapi

ALIGNMENTS

RESULT 1  
AC114884  
LOCUS  
DEFINITION  
AC114884  
144869 bp DNA linear HTG 25-SEP-2002  
Rattus norvegicus clone RP31-557J15 strain Brown Norway, WORKING  
DRAFT SEQUENCE, 8 ordered pieces.  
ACCESSION  
AC114884.2 GI:23308021  
VERSION  
HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
KEYWORDS  
Rattus norvegicus (Norway rat)  
SOURCE  
Rattus norvegicus  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE  
1 (bases 1 to 144869)  
Akhter N., Antonellis A., Ayele K., Beckstrom-Sternberg S.M.,

Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
 Cariaga, K., Coleman, B., Dietrich, N.L., Granite, S., Guan, X.,  
 Gupta, J., Raghigghi, P., Han, J., Hansen, N., Ho, S.-L., Idol, J.R.,  
 Karlins, E., Laric, P., Lee-Lin, S.-Q., Legaspi, R., Maduro, Q.L.,  
 Maduro, V.B., Margulies, E.H., Masiello, C., Maskeri, B.,  
 Mastrian, S.D., McCloskey, J.C., McDowell, J., Paquirigan, C.,  
 Pearson, R., Portnoy, M.E., Prasad, A., Reddix-Dugue, N.,  
 Schueler, M.G., Sison, C., Stantripop, S., Thomas, J.W., Thomas, P.J.,  
 Touchman, J.W., Vogt, J.L., Walker, M., Wetherby, K.D., Wiggins, L.,  
 Young, A., Zhang, L.-H., and Green, E.D.  
 NISC Comparative Sequencing Initiative  
 Unpublished  
 2 (bases 1 to 144869)  
 Direct Submission  
 Submitted (13-MAR-2002) NIH Intramural Sequencing Center, 8717  
 Government Circle, Gaithersburg, MD 20877, USA  
 3 (bases 1 to 144869)  
 Green, E.D.  
 Direct Submission  
 Submitted (25-SEP-2002) NIH Intramural Sequencing Center, 8717  
 Government Circle, Gaithersburg, MD 20877, USA  
 On Sep 25, 2002 this sequence version replaced gi:19387611.  
 ----- Genome Center  
 Center: NIH Intramural Sequencing Center  
 Center code: NISC  
 Web site: <http://www.nisc.nih.gov>  
 Contact: [nisc.zoehgri.nih.gov](mailto:nisc.zoehgri.nih.gov)  
 ----- Project Information  
 Center project name: cxn  
 Center clone name: 557J35

The sequence data in this record represents an 'enhanced'  
 version of a Phase 2 submission. Specifically, the indicated  
 order and orientation of each sequence contig has been  
 established using one or more of the following: read-pair  
 data from individual subclones, overlaps with neighboring  
 clones, alignment with available reference sequence (e.g.,  
 human), and/or confirmation by PCR testing. In addition,  
 the sequence assembly is based on at least 8x average  
 coverage in Q20 bases and has been reviewed to rule out  
 gross misassemblies, the low-quality ends of sequence  
 contigs have been trimmed away, and each base is associated  
 with a Phrap derived quality score.

----- Summary Statistics  
 Sequencing vector: plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 143282 bases at least Q40  
 Consensus quality: 143743 bases at least Q30  
 Consensus quality: 144038 bases at least Q20  
 Insert size: 151000; agarose-fp  
 Insert size: 144169; sum-of-contigs  
 Quality coverage: 10.5ix in Q20 bases; agarose-fp  
 Quality coverage: 11.0ix in Q20 bases; sum-of-contigs

-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* been provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 48966: contig of 48966 bp in length  
 \* 49067 49066: gap of unknown length  
 \* 49067 74022: contig of 24956 bp in length  
 \* 74023 74122: gap of unknown length  
 \* 74123 76637: contig of 2515 bp in length  
 \* 76638 76737: gap of unknown length  
 \* 76738 82524: contig of 5787 bp in length  
 \* 82525 82624: gap of unknown length

\* 82625 93442: contig of 16718 bp in length  
 \* 93443 93442: gap of unknown length  
 \* 93443 110906: contig of 11464 bp in length  
 \* 110907 110906: gap of unknown length  
 \* 110907 121200: contig of 10194 bp in length  
 \* 121201 121300: gap of unknown length  
 \* 121301 144869: contig of 23569 bp in length.  
 FEATURES  
 Location/Qualifiers  
 1. 144869  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /strain="Brown Norway"  
 /db\_xref="taxon:10116"  
 /clone="RP31-557J15"  
 /clone\_lib="RP31"  
 1. 59406  
 /note="clone overlaps with GenBank Accession Number  
 AC114881 clone RP31-385022 (center project name cxo)"  
 1. 48966  
 /note="assembly\_fragment  
 clone\_end:T7  
 vector\_side:left"  
 49067..74022  
 /note="assembly\_fragment"  
 74123..76637  
 /note="assembly\_fragment"  
 76738..82524  
 /note="assembly\_fragment"  
 82625..93442  
 /note="assembly\_fragment"  
 93443..110906  
 /note="assembly\_fragment"  
 110907..121200  
 /note="assembly\_fragment"  
 121301..144869  
 /note="assembly\_fragment  
 clone\_end:SP5  
 vector\_side:right"  
 39556 a 35181 c 34455 g 34977 t 700 others  
 BASE COUNT  
 ORIGIN  
 Query Match 100.0%; Score 18; DB 2; Length 144869;  
 Best Local Similarity 100.0%; Pred. No. 67;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GACAGGGCAGCAGACAG 18  
 |||||  
 Db 38338 GACAGGGCAGCAGACAG 38355  
 RESULT 2  
 AC114881  
 LOCUS  
 DEFINITION Rattus norvegicus clone RP31-385022 strain Brown Norway, WORKING  
 DRAFT SEQUENCES, 4 ordered pieces.  
 ACCESSION AC114881  
 VERSION AC114881.2 GI:21629279  
 KEYWORDS HTGS, PHASE2; HTGS, DRAFT.  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 REFERENCE  
 1 (bases 1 to 178857)  
 AUTHORS Akhter, N., Antonellis, A., Avele, K., Beckstrom-Sternberg, S.M.,  
 Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C.,  
 Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J.,  
 Haghighi, P., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E., Laric, P.,  
 Lee-Lin, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.B.,  
 Margulies, E.H., Masiello, C., Maskeri, B., Mastrian, S.D.,  
 McCloskey, J.C., McDowell, J., Paquirigan, C., Pearson, R.,  
 Portnoy, M.E., Prasad, A., Schueler, M.G., Stantripop, S., Thomas, J.W.,  
 Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,

Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.  
 NISC Comparative Sequencing Initiative  
 Unpublished  
 2 (bases 1 to 178857)  
 Green, E.D.  
 Direct Submission  
 Submitted (13-MAR-2003) NIH Intramural Sequencing Center, 8717  
 Grovmont Circle, Gaithersburg, MD 20877, USA  
 3 (bases 1 to 178857)  
 Green, E.D.  
 Direct Submission  
 Submitted (29-JUN-2002) NIH Intramural Sequencing Center, 8717  
 Grovmont Circle, Gaithersburg, MD 20877, USA  
 On Jun 29, 2002 this sequence version replaced gi:19387608.  
 -----  
 Genome Center  
 Center: NIH Intramural Sequencing Center  
 Center code: NISC  
 Web site: <http://www.nisc.nih.gov>  
 Contact: [nisc.zoo@ngri.nih.gov](mailto:nisc.zoo@ngri.nih.gov)  
 -----  
 Project Information  
 Center project name: czo  
 Center clone name: 385022

The sequence data in this record represents an 'enhanced' version of a phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics -----  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 178011 bases at least Q40  
 Consensus quality: 178364 bases at least Q30  
 Consensus quality: 178496 bases at least Q20  
 Insert size: 166000; agarose-fp  
 Insert size: 17857; sum-of-contigs  
 Quality coverage: 11.42x in Q20 bases; agarose-fp  
 Quality coverage: 10.61x in Q20 bases; sum-of-contigs

-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 4 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submittor.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

1 97908: contig of 97908 bp in length  
 97909 98008: gap of unknown length  
 98009 141249: contig of 43241 bp in length  
 141250 141349: gap of unknown length  
 141350 168491: contig of 27142 bp in length  
 168492 168591: gap of unknown length  
 168592 178857: contig of 10266 bp in length.

----- Location/Qualifiers -----

1. 178857  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /strain="Brown Norway"  
 /db\_xref="taxon:10116"  
 /clone\_lib="RP31-385022"  
 /clone\_lib="RP31"  
 1. 97908  
 /note="assembly\_fragment"

FEATURES  
 source

clone\_end:SP6  
 vector\_side:left"  
 98009..141249  
 /note="assembly\_fragment"  
 119530..178857  
 /note="clone overlaps with GenBank Accession Number  
 AC114884 clone RP31-557J15 (center project name cxn) "  
 141350..168491  
 /note="assembly\_fragment"  
 168592..178857  
 /note="assembly\_fragment"  
 clone\_end:T7  
 vector\_side:right"  
 47760 a 42335 c 42524 g 45938 t 300 others  
 BASE COUNT  
 ORIGIN  
 Query Match 100.0%; Score 18; DB 2; Length 178857;  
 Best Local Similarity 100.0%; Pred No. 65;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GACAAAGGCGACGACAGAG 18  
 |||||  
 Db 157801 GACAAAGGCGACGACAGAG 157818

RESULT 3  
 AC118733/c  
 LOCUS  
 DEFINITION  
 AC118733 188490 bp DNA linear HTG 22-FEB-2003  
 Mus musculus clone RP24-170A20, WORKING DRAFT SEQUENCE, 3 unordered  
 pieces.  
 AC118733 3 GI:28467248  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 Mus musculus (house mouse)  
 SOURCE  
 Mus musculus  
 ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 188490)  
 Birren,B., Birren,B., Nusbaum,C. and Lander,E.  
 Mus musculus, clone RP24-170A20  
 Unpublished  
 2 (bases 1 to 188490)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,D.,  
 Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
 Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kanat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R.,  
 Landers,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
 McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,  
 Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
 Oliver,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
 Olier,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (20-APR-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 188490)

JOURNAL

AUTHORS

AUTHORS

AUTHORS





Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIR

Web site: <http://www.seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L17758

Center clone name: 63\_O\_23

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

\* This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 209706: contig of 209706 bp in length.

#### FEATURES

source

Location/Qualifiers  
1. 209706  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="RP24-63023"  
/clone\_lib="RBC1-24 Male Mouse BAC"

65611 a 41280 c 40670 g 62145 t

BASE COUNT  
ORIGIN

Query Match 100.0%; Score 18; DB 2; Length 209706;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GACAAGGGCAGGACAGAG 18

Db 32614 GACAAGGGCAGGACAGAG 32631

#### RESULT 6

AC103052/c

LOCUS Rattus norvegicus clone CH230-147P15, WORKING DRAFT SEQUENCE.  
AC103052

AC103052.5 GI:30580047

HTG: HTGS\_PHASE2: HTGS\_DRAFT: HTGS\_FULLTOP.

Rattus norvegicus (Norway rat)

SOURCE

ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

1 (bases 1 to 217691)

#### REFERENCE

AUTHORS

Muzny D, Marle, Metzker M, Lee, Abramson S., Adams C., Alder J., Allen C., Allen H., Alsbrooks S., Amin A., Angiano D., Anyalebechi V., Aoyagi A., Ayodeji M., Baca E., Baden H., Baldwin D., Bandaranaike D., Barber M., Barnstead M., Benamed F., Biswal N., Blair J., Blankenburg K., Blyth P., Brown M., Bryant N., Buhay C., Burch P., Burrell K., Calderon E., Cardenas V., Carter K., Cavazos I., Ceasar H., Center A., Chacko J., Chavez D., Chen G., Chen R., Chen Y., Chen Z., Chu J., Cleveland C., Cockrell R., Cox C., Coyle M., Cree A., D'Souza L., Davila M.L., Davis C., Davy Carroll L., De Anda C., Dederich D., Delgado O., Denson S., Deramo C., Ding Y., Dinh H., Divya K., Draper H., Dugan-Rocha S., Dunn A., Durbin K., Duval T., Eaves K., Egan A., Escotte M., Eugene C., Evans C.A., Falls T., Fan G., Fernandez S., Finley M., Flag N., Forbes L., Foster M., Foster P., Fraser C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M., Gebregorgis E., Geer K., Gill R., Grady M., Guerra W., Guevara W., Gunaratne P., Haaland W., Hamill C., Hamilton C., Hamilton K., Harvey Y., Havlak P., Hawes A., Henderson N., Hernandez J., Hernandez R., Hines S., Hladun S.L., Hodgson A., Hogue M., Hollins B., Howles S., Hulyk S., Hume J., Ttlebird D., Jackson A., Jackson J., Jacob L., Jiang H., Johnson B., Johnson R., Jolivet A., Karpaty S., Kelly S., Kelly S., Khan Z., King L., Kovar C., Kowis C., Kraft C.L., Lebow H., Levan J., Lewis L., Li Z., Liu J.,

Liu J., Liu W., Liu Y., London P., Longacre S., Lopez J., Lorusuwa L., Loulseg H., Lozado R.J., Lu X., Ma J., Maheshwari M., Mahindartine M., Mahmud M., Malloy K., Mangum A., Mangum B., Mapua P., Martin K., Martin R., Martinez E., Mawhinney S., McLeod M.P., McNeill T.Z., Meenen E., Milosavljevic A., Miner G., Minja E., Montemayor J., Moore S., Morgan M., Morris K., Morris S., Munidasa M., Murphy M., Nair L., Nankervis C., Neal D., Newton N., Nguyen N., Norris S., Nwakoileh O., Okwuonu G., Olarnpunsagoon A., Pal S., Parks K., Pasternak S., Paul H., Perez A., Perez L., Pfankoch C., Popper F., Poindexter A., Popovic D., Primus E., Pu L., Puzos M., Quiroz J., Rachlin E., Reeves K., Regier M.A., Reigh R., Reilly B., Reilly M., Ren Y., Reuter M., Richards S., Riggs F., Rivers C., Rokey T., Rojas A., Rose M., Rose R., Ruiz S.J., Sanders W., Savary G., Scherer S., Scott G., Shatsman S., Shen H., Shetty J., Shvartsbeyn A., Sisson I., Sitter C.D., Smajls D., Sneed A., Sodergren E., Song X.-Z., Sorelle R., Sosa J., Steimle M., Strong R., Sutton A., Svatek A., Tabor P., Taylor C., Taylor R., Thomas N., Thomas S., Tingey A., Trejos Z., Usmani K., Valas R., Vera V., Villadana D., Waldron L., Walker B., Wang J., Wang Q., Wang S., Warren J., Warren R., Wei X., White F., Williams G., Willson R., Wleczyk R., Wooden H., Worley K., Wright D., Wright R., Wu J., Yakub S., Yen J., Yoon L., Yoon V., Yu F., Zhang J., Zhou J., Zhou X., Zhao S., Dunn D., von Niederhausen A., Weiss R., Smith D.R., Holt R.A., Smith H.O., Weinstock G. and Gibbs R.A.

#### Direct Submission

Unpublished  
2 (bases 1 to 217691)

Worley K.C.

#### Direct Submission

Submitted (24-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 217691)

Rat Genome Sequencing Consortium.

#### Direct Submission

Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 13, 2003 this sequence version replaced gi:23614691.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GJCH

Center clone name: CH230-147P15

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 212164 bases at least Q40

Consensus quality: 213776 bases at least Q30

Consensus quality: 215303 bases at least Q20

Estimated insert size: 223883; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 1 contigs. Gaps between the contigs



\* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* this sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 217691: contig of 217691 bp in length.

## FEATURES

Location/Qualifiers  
 1. 217691  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-147P15"  
 misc\_feature  
 1. 1410  
 /note="wgs-contig"  
 BASE COUNT 54935 a 50714 c 51140 g 59960 t 942 others  
 ORIGIN

Query Match 100.0%; Score 18; DB 2; Length 217691;  
 Best Local Similarity 100.0%; Pred. No. 64;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACAGGCGAGCAGAG 18

Db 129742 GACAGGCGAGCAGAG 129725

## RESULT 7

AC119862/c  
 LOCUS AC119862 244620 bp DNA linear HTG 23-APR-2003  
 DEFINITION Mus musculus clone RP24-93G1, WORKING DRAFT SEQUENCE, 10 unordered  
 pieces.

AC119882

AC119882.3 GI:30017888

HTG: HTGS-PHASE1; HTGS DRAFT.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 244620)

Birren,B., Nusbaum,C. and Lander,E.

Mus musculus, clone RP24-93G1

Unpublished

2 (bases 1 to 244620)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,  
 Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Fard,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,  
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R.,  
 Landers,T., Lechoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
 McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,  
 Mihova,T., Milenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,  
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
 Raymond,C., Retta,R., Riback,M., Riley,R., Rise,C., Rogov,P.,  
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (02-MAY-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 244620)

## REFERENCE

Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
 Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,

## TITLE

## JOURNAL

## COMMENT

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L25016

Center clone name: 93\_G\_1

-----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 10 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 2154: contig of 2154 bp in length  
 \* 2155 2254: gap of 100 bp  
 \* 2255 2921: contig of 667 bp in length  
 \* 2922 3021: gap of 100 bp  
 \* 3022 3879: contig of 858 bp in length  
 \* 3880 3979: gap of 100 bp  
 \* 3980 4654: contig of 675 bp in length  
 \* 4655 4754: gap of 100 bp  
 \* 4755 13630: contig of 8876 bp in length  
 \* 13631 13730: gap of 100 bp  
 \* 13731 36298: contig of 22568 bp in length  
 \* 36299 36398: gap of 100 bp  
 \* 36399 65869: contig of 29471 bp in length  
 \* 65870 65969: gap of 100 bp  
 \* 65970 106402: contig of 40433 bp in length  
 \* 106403 106502: gap of 100 bp  
 \* 106503 184024: contig of 77522 bp in length  
 \* 184025 184124: gap of 100 bp  
 \* 184125 244620: contig of 60496 bp in length.

## FEATURES

## source

1. 244620  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /clone="RP24-93G1"  
 /clone\_lib="RPCI-24 Male Mouse BAC"

## misc\_feature

1. 2154  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:left"  
 2255..2921

## misc\_feature

misc\_feature /note="assembly\_fragment"  
3022.3879  
misc\_feature /note="assembly\_fragment"  
3980.4654  
misc\_feature /note="assembly\_fragment"  
4755.13630  
misc\_feature /note="assembly\_fragment"  
13731.36238  
misc\_feature /note="assembly\_fragment"  
36399.65869  
misc\_feature /note="assembly\_fragment"  
65970.106402  
misc\_feature /note="assembly\_fragment"  
106503.184024  
misc\_feature /note="assembly\_fragment"  
184125.244620  
misc\_feature /note="assembly\_fragment"

BASE COUNT 68543 a 53239 c 52884 g 69050 t 904 others  
ORIGIN

Query Match 100.0%; Score 18; DB 2; Length 244620;  
Best Local Similarity 100.0%; Fred. No. 63;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACAAGGCGACGACGAG 18

Db 46216 GACAAGGCGACGACGAG 46199

RESULT 8  
AC124053 47589 bp DNA linear HTG 28-AUG-2002  
LOCUS Mus musculus clone RP24-252H13, LOW-PASS SEQUENCE SAMPLING.  
AC124053  
ACCESSION AC124053  
VERSION AC124053.2 GI:22532607  
KEYWORDS HTG; HTGS; PHASE0.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 47589)  
AUTHORS Birren,B., Nusbaum,C. and Lander,E.  
TITLE Mus musculus, clone RP24-252H13  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 47589)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,  
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzgerald,M., Fitzhugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginder,S., Gord,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,  
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,  
Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N.,  
Matthews,C., McCarthy,M., McSwan,P., McKernan,K., Meldrum,J.,  
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,  
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,  
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,  
Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,  
Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission  
JOURNAL Submitted (09-JUN-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
REFERENCE 3 (bases 1 to 47589)  
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,  
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzgerald,M., Fitzhugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginder,S., Gord,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,  
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,  
Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N.,  
Matthews,C., McCarthy,M., McSwan,P., McKernan,K., Meldrum,J.,  
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,  
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,  
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,  
Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,  
Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,  
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,  
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mienga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,  
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (28-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Aug 28, 2002 this sequence version replaced gi:21360073.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIER  
Web site: http://www.seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L21231  
Center clone name: 252\_H13  
-----

\* NOTE: This record contains 59 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
637: contig of 637 bp in length  
\* 638 737: gap of 100 bp  
\* 738 1406: contig of 669 bp in length  
\* 1407 1506: gap of 100 bp  
\* 1507 2190: contig of 684 bp in length  
\* 2191 2290: gap of 100 bp  
\* 2291 3051: contig of 761 bp in length  
\* 3052 3151: gap of 100 bp  
\* 3152 3917: contig of 766 bp in length  
\* 3918 4017: gap of 100 bp  
\* 4018 4793: contig of 776 bp in length  
\* 4794 4893: gap of 100 bp  
\* 4894 5683: contig of 790 bp in length  
\* 5684 5783: gap of 100 bp  
\* 5784 6448: contig of 665 bp in length  
\* 6449 6548: gap of 100 bp  
\* 6549 7326: contig of 678 bp in length  
\* 7327 8092: contig of 766 bp in length  
\* 8093 8192: gap of 100 bp  
\* 8193 8951: contig of 759 bp in length  
\* 8952 9051: gap of 100 bp  
\* 9052 9710: contig of 659 bp in length  
\* 9711 9810: gap of 100 bp  
\* 9811 10479: contig of 669 bp in length  
\* 10480 10579: gap of 100 bp  
\* 10580 11246: contig of 667 bp in length  
\* 11247 11346: gap of 100 bp  
\* 11347 12018: contig of 672 bp in length  
\* 12019 12118: gap of 100 bp

\* 12119 12784: contig of 666 bp in length  
\* 12785 12884: gap of 100 bp  
\* 12885 13554: contig of 670 bp in length  
\* 13555 13654: gap of 100 bp  
\* 13655 14420: contig of 766 bp in length  
\* 14421 14520: gap of 100 bp  
\* 14521 15213: contig of 693 bp in length  
\* 15214 15313: gap of 100 bp  
\* 15314 15988: contig of 675 bp in length  
\* 15989 16088: gap of 100 bp  
\* 16089 16752: contig of 684 bp in length  
\* 16753 16852: gap of 100 bp  
\* 16853 17525: contig of 673 bp in length  
\* 17526 17625: gap of 100 bp  
\* 17626 18298: contig of 673 bp in length  
\* 18299 18398: gap of 100 bp  
\* 18399 19093: contig of 695 bp in length  
\* 19094 19193: gap of 100 bp  
\* 19194 19953: contig of 760 bp in length  
\* 19954 20053: gap of 100 bp  
\* 20054 20713: contig of 660 bp in length  
\* 20714 20813: gap of 100 bp  
\* 20814 21480: contig of 667 bp in length  
\* 21481 21580: gap of 100 bp  
\* 21581 22338: contig of 738 bp in length  
\* 22339 22438: gap of 100 bp  
\* 22439 23123: contig of 685 bp in length  
\* 23124 23223: gap of 100 bp  
\* 23224 23905: contig of 682 bp in length  
\* 23906 24005: gap of 100 bp  
\* 24006 24706: contig of 701 bp in length  
\* 24707 24805: gap of 100 bp  
\* 24806 25475: contig of 669 bp in length  
\* 25476 25575: gap of 100 bp  
\* 25576 26275: contig of 700 bp in length  
\* 26276 26375: gap of 100 bp  
\* 26376 27192: contig of 817 bp in length  
\* 27193 27292: gap of 100 bp  
\* 27293 28040: contig of 748 bp in length  
\* 28041 28140: gap of 100 bp  
\* 28141 28891: contig of 751 bp in length  
\* 28892 28991: gap of 100 bp  
\* 28992 29648: contig of 657 bp in length  
\* 29649 29748: gap of 100 bp  
\* 29749 30525: contig of 777 bp in length  
\* 30526 30625: gap of 100 bp  
\* 30626 31393: contig of 768 bp in length  
\* 31394 31493: gap of 100 bp  
\* 31494 32191: contig of 698 bp in length  
\* 32192 32291: gap of 100 bp  
\* 32292 32946: contig of 655 bp in length  
\* 32947 33046: gap of 100 bp  
\* 33047 33730: contig of 684 bp in length  
\* 33731 33830: gap of 100 bp  
\* 33831 34502: contig of 672 bp in length  
\* 34503 34602: gap of 100 bp  
\* 34603 35382: contig of 780 bp in length  
\* 35383 35482: gap of 100 bp  
\* 35483 36154: contig of 672 bp in length  
\* 36155 36254: gap of 100 bp  
\* 36255 36951: contig of 697 bp in length  
\* 36952 37051: gap of 100 bp  
\* 37052 37724: contig of 673 bp in length  
\* 37725 37824: gap of 100 bp  
\* 37826 38495: contig of 671 bp in length  
\* 38496 38595: gap of 100 bp  
\* 38596 39388: contig of 793 bp in length  
\* 39389 39488: gap of 100 bp  
\* 39489 40277: contig of 789 bp in length  
\* 40278 40377: gap of 100 bp  
\* 40378 41141: contig of 764 bp in length  
\* 41142 41241: gap of 100 bp  
\* 41242 41909: contig of 668 bp in length

\* 41910 42009: gap of 100 bp  
\* 42010 42696: contig of 687 bp in length  
\* 42697 42796: gap of 100 bp  
\* 42797 43463: contig of 667 bp in length  
\* 43464 43563: gap of 100 bp  
\* 43564 44224: contig of 661 bp in length  
\* 44225 44324: gap of 100 bp  
\* 44325 44985: contig of 661 bp in length  
\* 44986 45085: gap of 100 bp  
\* 45086 45862: contig of 777 bp in length  
\* 45863 45962: gap of 100 bp  
\* 45964 46732: contig of 770 bp in length

Query Match 94.4% Score 17; DB 2; Length 47589;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACAAGGCAGGACAGA 17  
|||||  
Db 36306 GACAAGGCAGGACAGA 36322

RESULT 9  
AC121356 116608 bp DNA linear HTG 16-MAR-2003  
LOCUS Mus musculus clone RP23-52D17, WORKING DRAFT SEQUENCE, 5 unordered  
DEFINITION pieces.  
ACCESSION AC121356  
VERSION AC121356.3 GI:28975044  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 116608)  
AUTHORS Birren,B., Nusbaum,C. and Lander,E.  
TITLE Mus musculus, clone RP23-52D17  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 116608)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,  
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzgerald,M., FitzGerald,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,  
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Lacroque,K.,  
Lamazares,R., Landers,T., Lenoczky,J., Levine,R., Lindblad-Toh,K.,  
Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,  
Matthews,C., McCarthy,M., McEwan,F., McKernan,K., Meldrum,J.,  
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
Pollara,V., Raymond,C., Retta,R., Rieback,W., Riley,R., Rise,C.,  
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,  
Shogback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,  
Theodore,J., Topham,X., Travers,M., Travis,N., Trigilio,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
TITLE Direct Submission  
JOURNAL  
REFERENCE 3 (bases 1 to 116608)  
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,  
Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,  
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,  
Graham,L., Grand-Pierre,N., Hagos,B., Hulme,W., Iliev,I.,  
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,

Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrum,J., Menes,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,I., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (16-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 16, 2003 this sequence version replaced gi:25901114.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www.seq.wi.mit.edu  
Contact: sequence.submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L23269  
Center clone name: 52\_D\_17  
----- Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 115769 bases at least Q40  
Consensus quality: 115995 bases at least Q30  
Consensus quality: 116130 bases at least Q20  
Insert size: 150000; agarose-fp  
Insert size: 116208; sum-of-contigs  
Quality coverage: 10.2 in Q20 bases; agarose-fp  
Quality coverage: 13.1 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 23731: contig of 23731 bp in length  
23732 23831: gap of 100 bp  
23832 41019: contig of 17188 bp in length  
41020 41119: gap of 100 bp  
41120 60788: contig of 19569 bp in length  
60789 60888: gap of 100 bp  
60889 97152: contig of 36264 bp in length  
97153 97252: gap of 100 bp  
97253 116608: contig of 19356 bp in length.

-----  
Location/Qualifiers  
1. .116608  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="RP23-52D17"  
/genome\_lib="RPC1-23 Female Mouse BAC"  
1. .23731  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:left"  
23832. .41019  
/note="assembly\_fragment"  
41120. .60788  
/note="assembly\_fragment"  
60889. .97152  
/note="assembly\_fragment"

misc\_feature  
source  
1. .116608  
/note="assembly\_fragment  
clone\_end:SP6  
vector\_side:right"  
BASE COUNT 34865 a 23112 c 22833 g 35397 t 401 others  
ORIGIN

Query Match 94.4%; Score 17; DB 2; Length 116608;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GACAGGCGCAGGACAGA 17  
|||||  
Db 36482 GACAGGCGCAGGACAGA 36498  
|||||

RESULT 10  
AC144886/c  
LOCUS  
DEFINITION  
AC144886  
AC144886.1 GI:31044305  
VERSION  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS  
SOURCE  
Bos taurus (cow)  
ORGANISM  
Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE  
1 (bases 1 to 151613)  
ANTONELLIS,A., AYELE,K., BECKSTROM-STERBERG,S.M., BENJAMIN,B.,  
BLAKESLEY,R.W., BOUFFARD,G.G., BRINKLEY,C., BROOKS,S., CARIAGA,K.,  
CHU,G., COLEMAN,B., COLEMAN,H., ENGLE,J., GRANITE,S., GUAN,X.,  
GUPTA,B., HAGHIGHI,P., HAN,J., HANSEN,N., HO,S.-L., HU,P.,  
HURLE,B., IDOL,J.R., KARLINS,E., KWONG,P., LARIC,P., LEE-LIN,S.-Q.,  
LEGASPI,R., MADURO,O.L., MADURO,V.B., MARGULIES,E.H., MASIELLO,C.,  
MASKERI,B., MCDOWELL,J., PAGUIRICAN,C., PEARSON,R., PORTNOV,M.E.,  
PRAAD,A., REDDIX-DUGUE,N., SCHANDLER,K., SCHUELER,M.G., SHAH,K.,  
SISON,C., STANTIPOP,S., THOMAS,J.W., THOMAS,P.J., TSIPOURI,V.,  
VOGT,J.L., WETHERBY,K.D., WIGGINS,L., YOUNG,A. and GREEN,E.D.

-----  
TITLE  
JOURNAL  
REFERENCE  
2 (bases 1 to 151613)  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Submitted (23-MAY-2003) NIH Intramural Sequencing Center, 8717  
Grovefront Circle, Galtersburg, MD 20877, USA  
-----  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: http://www.nisc.nih.gov  
Contact: nisc\_zoo@hgr.nih.gov  
-----  
Project Information  
Center project name: djc  
Center clone name: 363J22  
-----  
Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 149535 bases at least Q40  
Consensus quality: 146470 bases at least Q30  
Consensus quality: 147291 bases at least Q20  
Insert size: 120000; agarose-fp  
Insert size: 150213; sum-of-contigs  
Quality coverage: 11.18x in Q20 bases; agarose-fp  
Quality coverage: 8.93x in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will  
\* be preserved.

```

1      2844: contig of 2844 bp in length
*      2845: gap of unknown length
*      2945: contig of 2180 bp in length
*      5124: contig of 2180 bp in length
*      5125: gap of unknown length
*      5225: contig of 4536 bp in length
*      9760: gap of unknown length
*      9761: contig of 3211 bp in length
*      9861: contig of 3211 bp in length
*      13072: gap of unknown length
*      13171: gap of unknown length
*      13172: contig of 4621 bp in length
*      17792: gap of unknown length
*      17793: gap of unknown length
*      17893: contig of 6224 bp in length
*      24117: gap of unknown length
*      24217: contig of 6313 bp in length
*      30530: gap of unknown length
*      30629: gap of unknown length
*      30630: contig of 11438 bp in length
*      42057: gap of unknown length
*      42168: contig of 9408 bp in length
*      51575: gap of unknown length
*      51676: gap of unknown length
*      51676: contig of 10654 bp in length
*      62330: gap of unknown length
*      62430: contig of 10420 bp in length
*      72850: gap of unknown length
*      72850: contig of 13414 bp in length
*      86364: gap of unknown length
*      86464: contig of 13532 bp in length
*      99996: gap of unknown length
*      100096: contig of 22610 bp in length
*      122705: gap of unknown length
*      122805: contig of 28808 bp in length.
*      122806: Location/Qualifiers
1. 151613
  /organism="Bos taurus"
  /mol_type="genomic DNA"
  /db_xref="taxon:9913"
  /clone="RP42-363J22"
  /clone_lib="RP42"

```

#### FEATURES

source

```

misc_feature   1. 2844
               /note="assembly_fragment"
misc_feature   2945. 5124
               /note="assembly_fragment"
misc_feature   5225. 9760
               /note="assembly_fragment"
misc_feature   9861. 13071
               /note="assembly_fragment"
misc_feature   13172. 17792
               /note="assembly_fragment"
misc_feature   17893. 24116
               /note="assembly_fragment"
misc_feature   24217. 30529
               /note="assembly_fragment"
misc_feature   30630. 42067
               /note="assembly_fragment"
misc_feature   42168. 51575
               /note="assembly_fragment"
misc_feature   51676. 62329
               /note="assembly_fragment"
misc_feature   62430. 72849
               /note="assembly_fragment"
misc_feature   72950. 86363
               /note="assembly_fragment"
misc_feature   86464. 99995
               /note="assembly_fragment
               clone_end:SP6
               vector_side:right"
misc_feature   100096. 122705
               /note="assembly_fragment"
misc_feature   122806. 151613
               /note="assembly_fragment
               clone_end:T7
               vector_side:right"

```

BASE COUNT 31102 a 44763 c 42876 g 31470 t 1402 others

#### ORIGIN

```

Query Match      94.4%; Score 17; DB 2; Length 151613;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GACAAGGCGAGGACAGA 17
      |||||
Db      43118 GACAAGGCGAGGACAGA 43102

RESULT 11
LOCUS      AC026761                163475 bp    DNA    linear    ROD 21-SEP-2002
DEFINITION Mus musculus Strain 129S6/SvEvTac chromosome 19 BAC, RP22-325J22,
            Complete sequence, complete sequence.
ACCESSION  AC026761
VERSION     AC026761.17    GI:22213163
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 163475)
AUTHORS     Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
            Toshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
            Gordon,M., Goltz,J.S. and Kucherlapati,R.
TITLE       Mouse High Throughput Sequencing
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 163475)
AUTHORS     Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
            Toshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
            Gordon,M., Goltz,J.S. and Kucherlapati,R.
TITLE       Direct Submission
JOURNAL     Submitted (24-MAR-2000) Department of Molecular Genetics, Albert
            Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
            Bronx, NY 10463, USA
REFERENCE   3 (bases 1 to 163475)
AUTHORS     Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
            Toshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
            Gordon,M., Goltz,J.S. and Kucherlapati,R.
TITLE       Direct Submission
JOURNAL     Submitted (29-JUN-2002) Harvard Partners Center for Genetics and
            Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
            02139, USA
REFERENCE   4 (bases 1 to 163475)
AUTHORS     Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
            Toshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
            Gordon,M., Goltz,J.S. and Kucherlapati,R.
TITLE       Direct Submission
JOURNAL     Submitted (13-AUG-2002) Harvard Partners Center for Genetics and
            Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
            02139, USA
REFERENCE   5 (bases 1 to 163475)
AUTHORS     Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
            Toshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
            Gordon,M., Goltz,J.S. and Kucherlapati,R.
TITLE       Direct Submission
JOURNAL     Submitted (21-SEP-2002) Harvard Partners Center for Genetics and
            Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
            02139, USA
COMMENT     On Aug 13, 2002 this sequence version replaced gi:21629248.
            -----Genome Center:
            Center: Harvard Partners Genome Center
            Center Code: HPGC
            Web site: http://www.hpcgg.org/Sequence/mouse.html
            Contact: hpgc@mcendel.mgh.harvard.edu

```

CLONE LENGTH: This sequence represents the entire insert of this clone unless otherwise noted. If there are overlapping clones, the overlaps are noted in the beginning and end of the Features listing.



```

Db      125435  ACAAGGGCAGGACAGAG 125451
|||||
AC145188      173362 bp  DNA  linear  HTG 12-JUN-2003
Bos taurus clone RP42-483023, WORKING DRAFT SEQUENCE, 16 unordered
pieces.
ACCESSION  AC145188
VERSION    AC145188.1  GI:31621331
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE    Bos taurus (cow)
ORGANISM  Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 173362)
Antonelias,A., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Cariaga,K.,
Chu,G., Coleman,B., Coleman,H., Engle,J., Granite,S., Guan,X.,
Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,
Hurle,B., Idol,J.R., Karlins,E., Kwong,P., Laric,P., Lee-Lin,S.-Q.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Mastiello,C.,
Maskeri,B., McDowell,J., Paguirigan,C., Pearson,R., Portnoy,M.E.,
Prasad,A., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Shah,K.,
Sison,C., Stantripop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V.,
Vost,J.L., Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 173362)
Green,E.D.
Direct Submission
Submitted (12-JUN-2003) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
Center project name: djd
Center clone name: 483023
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 168024 bases at least Q40
Consensus quality: 169046 bases at least Q30
Consensus quality: 169811 bases at least Q20
Insert size: 158000; agarose-fp
Insert size: 171862; sum-of-contigs
Quality coverage: 11.06x in Q20 bases; agarose-fp
Quality coverage: 10.17x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2487: contig of 2487 bp in length
* 2488 2587: gap of unknown length
* 2588 4812: contig of 2225 bp in length
* 4813 4912: gap of unknown length
* 4913 9311: contig of 4399 bp in length
* 9312 9411: gap of unknown length
* 9412 11916: contig of 2505 bp in length
* 11917 12017: gap of unknown length
* 12017 14880: contig of 2864 bp in length
* 14881 14980: gap of unknown length
*
* 14981 18347: contig of 3367 bp in length
* 18348 18447: gap of unknown length
* 18448 26843: contig of 8396 bp in length
* 26844 26943: gap of unknown length
* 26944 33493: contig of 6556 bp in length
* 33494 33593: gap of unknown length
* 33594 43191: contig of 9592 bp in length
* 43192 43291: gap of unknown length
* 43292 55753: contig of 12462 bp in length
* 55754 67920: contig of 12067 bp in length
* 67921 79042: gap of unknown length
* 79043 79142: gap of unknown length
* 79143 93788: contig of 14646 bp in length
* 93789 107938: contig of 14050 bp in length
* 107939 123823: contig of 17787 bp in length
* 123824 125925: gap of unknown length
* 125926 173362: contig of 47437 bp in length.
*
FEATURES             Location/Qualifiers
     misc_feature     1..2487
                        /note="assembly_fragment"
     misc_feature     2588..4812
                        /note="assembly_fragment"
     misc_feature     4913..9311
                        /note="assembly_fragment"
     misc_feature     9412..11916
                        /note="assembly_fragment"
     misc_feature     12017..14880
                        /note="assembly_fragment"
     misc_feature     14981..18347
                        /note="assembly_fragment"
     misc_feature     18448..26843
                        /note="assembly_fragment"
     misc_feature     26944..33493
                        /note="assembly_fragment"
     misc_feature     33600..43191
                        /note="assembly_fragment"
     misc_feature     43292..55753
                        /note="assembly_fragment"
     misc_feature     55854..67920
                        /note="assembly_fragment"
     misc_feature     68021..79042
                        /note="assembly_fragment"
     misc_feature     79143..93788
                        /note="assembly_fragment"
     misc_feature     93889..107938
                        /note="assembly_fragment"
     misc_feature     108039..125825
                        /note="assembly_fragment"
     misc_feature     125926..173362
                        /note="assembly_fragment"
BASE COUNT  33233 a 52193 c 51596 g 34827 t 1513 others
ORIGIN
Query Match      94.4%; Score 17; DB 2; Length 173362;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 GACAAAGGGCAGGACAGAGA 17
|||||
Db      35764  GACAAAGGGCAGGACAGAGA 35748

```

```

RESULT 13
AC138402/c
LOCUS
DEFINITION
Mus musculus chromosome 14 clone RP24-263011 map 14, WORKING DRAFT
SEQUENCE, 3 unordered pieces.
AC138402
ACCESSION
AC138402.3 GI:28876093
VERSION
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 196351)
AUTHORS
Birren,B., Nusbaum,C. and Lander,E.
TITLE
Mus musculus chromosome 14, clone RP24-263011
JOURNAL
Unpublished
AUTHORS
2 (bases 1 to 196351)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Illiev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meidrim,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Willson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (29-DEC-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 196351)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cooke,P., Cooke,P., Corum,B., Dearellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,I., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Illiev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meidrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Willson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (07-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 7, 2003 this sequence version replaced gi:28461005.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RN/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
-----

Center project name: L28846
Center clone name: 263_0.11
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 195646 bases at least Q40
Consensus quality: 196011 bases at least Q30
Consensus quality: 196126 bases at least Q20
Insert size: 192000; agarose-fp
Insert size: 196151; sum-of-contigs
Quality coverage: 12.1 in Q20 bases; agarose-fp
Quality coverage: 11.8 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 11993: contig of 11993 bp in length
* 11994 12093: gap of 100 bp
* 12094 76435: contig of 84342 bp in length
* 76436 76535: gap of 100 bp
* 76536 196351: contig of 119816 bp in length.
* Location/Qualifiers
* 1. 196351
* /organism="Mus musculus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10090"
* /chromosome="14"
* /map="14"
* /clone="RP24-263011"
* /clone_lib="RPCI-24 Male Mouse BAC"
* 1. 11993
* /note="assembly_fragment
* clone_end:SP6
* vector_side:left"
* 12094..76435
* /note="assembly_fragment"
* 76536..196351
* /note="assembly_fragment
* clone_end:T7
* vector_side:right"
* BASE COUNT 56384 a 40543 c 40822 g 58402 t 200 others
* ORIGIN
*
* Query Match 94.4%; Score 17; DB 2; Length 196351;
* Best Local Similarity 100.0%; Pred.No. 2.2e+02;
* Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*
* Qy 1 GACAAAGGCGCAGGACAGA 17
* |||||
* Db 142339 GACAAAGGCGCAGGACAGA 142323
*
* RESULT 14
* AC079082
* LOCUS
* DEFINITION
* Mus musculus chromosome 10 clone rp23-161b11 strain C57BL/6J,
* complete sequence.
* ACCESSION
* AC079082
* VERSION
* AC079082.39 GI:25013332
* KEYWORDS
* HTG.
* SOURCE
* Mus musculus (house mouse)
* ORGANISM
* Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
* Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
* YING,F. and Roe,B.A.
* TITLE
* Mus musculus BAC Clone rp23-161b11

```



```

JOURNAL
REFERENCE
2 (bases 1 to 205621)
AUTHORS
Ying.F. and Roe.B.A.
TITLE
Direct Submission
JOURNAL
Submitted (18-AUG-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
3 (bases 1 to 205621)
AUTHORS
Ying.F. and Roe.B.A.
TITLE
Direct Submission
JOURNAL
Submitted (04-MAY-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
4 (bases 1 to 205621)
AUTHORS
Ying.F. and Roe.B.A.
TITLE
Direct Submission
JOURNAL
Submitted (07-MAY-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
5 (bases 1 to 205621)
AUTHORS
Ying.F. and Roe.B.A.
TITLE
Direct Submission
JOURNAL
Submitted (15-NOV-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
On Nov 15, 2002 this sequence version replaced gi:20429347.
-----
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
FEATURES
source
1. 205621
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/chromosome="10"
/clone="fp23-161b11"
/clone_lib="RCCI - 23 Female (C57BL/6J) Mouse BAC Library"
BASE COUNT 52129 a 53223 c 50420 g 49849 t
ORIGIN
Query Match 94.4%; Score 17; DB 10; Length 205621;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 ACAAGGCAGGACAGAG 18
Db 16542 ACAAGGCAGGACAGAG 16558
RESULT 15
AC132033
LOCUS
DEFINITION
Rattus norvegicus clone CH230-127G19, WORKING DRAFT SEQUENCE, 3
unordered pieces.
ACCESSION
AC132033
VERSION
AC132033.4 GI:25091414
KEYWORDS
HTG; HTGS_PRAISEI; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus
1 (bases 1 to 217181)
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Ayoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Devila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Georgescu,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoyues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpaty,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowitz,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulsegod,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Morris,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfandkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczek,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
-----
TITLE
Unpublished
JOURNAL
2 (bases 1 to 217181)
REFERENCE
Rat Genome Sequencing Consortium.
AUTHORS
Direct Submission
TITLE
Submitted (29-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
JOURNAL
3 (bases 1 to 217181)
REFERENCE
Rat Genome Sequencing Consortium.
AUTHORS
Direct Submission
TITLE
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
JOURNAL
On Nov 19, 2002 this sequence version replaced gi:22856251.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
-----
Center: Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/

```

```

Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GWSX
Center clone name: CH230-127G19
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 212810 bases at least Q40
Consensus quality: 213651 bases at least Q30
Consensus quality: 214227 bases at least Q20
Estimated insert size: 218471; sum-of-contigs estimation
Quality coverage: 9x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 211237: contig of 211237 bp in length
* 211238 211337: gap of unknown length
* 211338 214612: contig of 3275 bp in length
* 214613 214712: gap of unknown length
* 214713 217181: contig of 2469 bp in length.
*
* Location/Qualifiers
*   1. 217181
*     /organism="Rattus norvegicus"
*     /mol_type="genomic DNA"
*     /db_xref="taxon:10116"
*     /clone="CH230-127G19"
*
* BASE COUNT  60118 a 46007 c 47603 g 60919 t  2534 others
* ORIGIN
*
* Query Match      94.4%; Score 17; DB 2; Length 217181;
* Best Local Similarity 100.0%; Pred. No. 2.2e+02;
* Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*
* QY      2 ACAGGGCAGGACAGAG 18
*         |||||
* Db      12734 ACAGGGCAGGACAGAG 12750
*
* Search completed: August 27, 2003, 18:41:51
* Job time : 542.538 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 27, 2003, 18:42:00 ; Search time 173.269 Seconds  
(without alignments)

450.193 Million cell updates/sec

Title: US-10-085-612-2

Perfect score: 34

Sequence: 1 cgatttttctactgctgcagctgcagccccc 34

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1533700 seqs, 1147125425 residues

Total number of hits satisfying chosen parameters: 3067400

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	100.0	34	14	US-10-085-612-2
2	33	97.1	34	14	US-10-085-612-6
3	33	97.1	1254	14	US-10-085-612-4
4	22.8	67.1	1012	10	US-09-957-997-4
5	22.8	67.1	1345	14	US-10-146-575-3
6	22.8	67.1	1345	14	US-10-085-612-3
7	22.8	67.1	11186	10	US-09-957-997-1
8	22.8	67.1	13035	12	US-10-121-960C-14
9	22.8	67.1	15185	12	US-10-121-960C-17
10	21.6	63.5	481	13	US-10-027-632-72580
c 11	21.4	62.9	599	13	US-10-027-632-285947
c 12	20.6	60.6	30	14	US-10-215-432-41
c 13	20.2	59.4	470	11	US-09-918-995-24594
14	20.2	59.4	672	13	US-10-027-632-253693
15	20.2	59.4	994	11	US-09-934-455-489
16	20	58.8	627	9	US-09-770-149-726

Sequence 322, App  
Sequence 11, Appl  
Sequence 204795,  
Sequence 3264, Ap  
Sequence 3, Appl  
Sequence 1, Appl  
Sequence 1908, Ap  
Sequence 3727, Ap  
Sequence 31, Appl  
Sequence 278, App  
Sequence 255971,  
Sequence 255972,  
Sequence 158, App  
Sequence 334, App  
Sequence 24824, A  
Sequence 29796, A  
Sequence 8079, Ap  
Sequence 7049, Ap  
Sequence 50, Appl  
Sequence 570, Appl  
Sequence 513, App  
Sequence 26, Appl  
Sequence 3, Appl  
Sequence 486, App  
Sequence 5486, Ap  
Sequence 3210, Ap  
Sequence 4985, Ap  
Sequence 322496,

20 58.8 802 9 US-09-925-301-322  
20 58.8 1358 11 US-09-877-843-11  
19.8 58.2 1284 13 US-10-027-632-204795  
19.6 57.6 1485 14 US-10-156-761-3264  
19.6 57.6 249487 13 US-10-026-188-3  
19.6 57.6 9025608 14 US-10-156-761-1  
19.4 57.1 437 14 US-10-068-543-1908  
19.4 57.1 483 14 US-10-198-846-3727  
19.4 57.1 513 10 US-09-960-253-31  
19.4 57.1 697 9 US-09-770-149-278  
19.4 57.1 1573 13 US-10-027-632-255971  
19.4 57.1 1573 13 US-10-027-632-255972  
19.4 57.1 2838 10 US-09-960-253-158  
19.4 57.1 2838 14 US-10-171-581-334  
19.2 56.5 171 9 US-09-864-761-24824  
19.2 56.5 212 10 US-09-783-590-3651  
19.2 56.5 348 11 US-09-918-995-29798  
19.2 56.5 558 9 US-09-864-761-8079  
19.2 56.5 566 9 US-09-864-761-7049  
19.2 56.5 1763 14 US-10-177-293-50  
19.2 56.5 1516 9 US-09-925-301-570  
19.2 56.5 1936 10 US-09-925-300-513  
19.2 56.5 9980 11 US-09-998-716-26  
19.2 56.5 34185 14 US-10-199-520-3  
19 55.9 3103 14 US-10-128-714-486  
19 55.9 3103 14 US-10-128-714-486  
18.8 55.3 277 10 US-09-878-574-3210  
18.8 55.3 543 13 US-10-027-632-4985  
18.8 55.3 572 13 US-10-027-632-322496

#### ALIGNMENTS

#### RESULT 1

US-10-085-612-2  
; Sequence 2, Application US/10085612  
; Publication No. US20030096251A1  
; GENERAL INFORMATION:  
; APPLICANT: Guida, Marco  
; APPLICANT: Hall, Jeff  
; APPLICANT: Petros, William  
; APPLICANT: Vredenburgh, James  
; APPLICANT: Colvin, Oliver  
; APPLICANT: Marks, Jeffrey  
; TITLE OF INVENTION: Methods for Evaluating the Ability to Metabolize Pharmaceuticals  
; TITLE OF INVENTION: Compositions Therefor  
; FILE REFERENCE: 4389-S-CI  
; CURRENT APPLICATION NUMBER: US/10/085,612  
; CURRENT FILING DATE: 2002-02-26  
; PRIOR APPLICATION NUMBER: 09/144,367  
; PRIOR FILING DATE: 1998-08-31  
; PRIOR APPLICATION NUMBER: 60/271,630  
; PRIOR FILING DATE: 2001-02-26  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 34  
; TYPE: DNA  
; ORGANISM: Homo sapiens

#### US-10-085-612-2

Query Match 100.0%; Score 34; DB 14; Length 34;  
Best Local Similarity 100.0%; Pred. No. 0.00013;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGATCTTCTCTACTGCTGCAGCTGCAGCCCG 34

Db 1 CGATCTTCTCTACTGCTGCAGCTGCAGCCCG 34

#### RESULT 2

US-10-085-612-6

```

; Sequence 6, Application US/10085612
; Publication No. US20030096251A1
; GENERAL INFORMATION:
; APPLICANT: Guida, Marco
; APPLICANT: Hall, Jeff
; APPLICANT: Petros, William
; APPLICANT: Vredenburg, James
; APPLICANT: Colvin, Oliver
; APPLICANT: Marks, Jeffrey
; TITLE OF INVENTION: Methods for Evaluating the Ability to Metabolize Pharmaceuticals
; TITLE OF INVENTION: Compositions Therefor
; FILE REFERENCE: 4389-5-C1
; CURRENT APPLICATION NUMBER: US/10/085,612
; PRIOR FILING DATE: 2002-02-26
; PRIOR APPLICATION NUMBER: 09/144,367
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: 60/271,630
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 34
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-085-612-6

Query Match          97.1%; Score 33; DB 14; Length 34;
Best Local Similarity 100.0%; Pred. No. 0.00034;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATTCTTGGCTACTGGCTGCAGCTGCAGCCCG 33
    |||||||
DB 1 CGATTCTTGGCTACTGGCTGCAGCTGCAGCCCG 33

RESULT 3
US-10-085-612-4
; Sequence 4, Application US/10085612
; Publication No. US20030096251A1
; GENERAL INFORMATION:
; APPLICANT: Guida, Marco
; APPLICANT: Hall, Jeff
; APPLICANT: Petros, William
; APPLICANT: Vredenburg, James
; APPLICANT: Colvin, Oliver
; APPLICANT: Marks, Jeffrey
; TITLE OF INVENTION: Methods for Evaluating the Ability to Metabolize Pharmaceuticals
; TITLE OF INVENTION: Compositions Therefor
; FILE REFERENCE: 4389-5-C1
; CURRENT APPLICATION NUMBER: US/10/085,612
; PRIOR FILING DATE: 2002-02-26
; PRIOR APPLICATION NUMBER: 09/144,367
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: 60/271,630
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-085-612-4

Query Match          97.1%; Score 33; DB 14; Length 1254;
Best Local Similarity 100.0%; Pred. No. 0.00049;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATTCTTGGCTACTGGCTGCAGCTGCAGCCCG 33
    |||||||
DB 1004 CGATTCTTGGCTACTGGCTGCAGCTGCAGCCCG 1036

RESULT 4

```

```

US-09-957-997-4
; Sequence 4, Application US/09957997
; Patent No. US20020150915A1
; GENERAL INFORMATION:
; APPLICANT: Berkenstam, Anders
; APPLICANT: Bertilsson, Gran
; APPLICANT: Blomquist, Patrik
; TITLE OF INVENTION: PROMOTER SEQUENCES
; FILE REFERENCE: 13425-046001
; CURRENT APPLICATION NUMBER: US/09/957,997
; CURRENT FILING DATE: 2001-09-21
; EARLIER APPLICATION NUMBER: SE 0003393-6
; EARLIER FILING DATE: 2000-09-22
; EARLIER APPLICATION NUMBER: 60/238,895
; EARLIER FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1012
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-957-997-4

Query Match          67.1%; Score 22.8; DB 10; Length 1012;
Best Local Similarity 92.3%; Pred. No. 7;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 9 TGCTACTGGCTGCAGCTGCAGCCCG 34
    |||||||
DB 947 TGCTACTGGCTGCAGCTGCAGCCCG 972

RESULT 5
US-10-146-575-3
; Sequence 3, Application US/10146575
; Publication No. US20030059800A1
; GENERAL INFORMATION:
; APPLICANT: Lichter, Jay
; APPLICANT: Guido, Marco
; TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4
; FILE REFERENCE: SEQ-12P
; CURRENT APPLICATION NUMBER: US/10/146,575
; CURRENT FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: US/09/144,367
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1345
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: Other
; LOCATION: (0)...(0)
US-10-146-575-3

Query Match          67.1%; Score 22.8; DB 14; Length 1345;
Best Local Similarity 92.3%; Pred. No. 7.2;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 9 TGCTACTGGCTGCAGCTGCAGCCCG 34
    |||||||
DB 1035 TGCTACTGGCTGCAGCTGCAGCCCG 1060

RESULT 6
US-10-085-612-3
; Sequence 3, Application US/10085612
; Publication No. US20030096251A1
; GENERAL INFORMATION:
; APPLICANT: Guida, Marco
; APPLICANT: Hall, Jeff
; APPLICANT: Petros, William

```

```
; APPLICANT: Vredenburgh, James
; APPLICANT: Colwin, Oliver
; APPLICANT: Marks, Jeffrey
; TITLE OF INVENTION: Methods for Evaluating the Ability to Metabolize Pharmaceuticals
; FILE REFERENCE: 4389-5-C1
; CURRENT APPLICATION NUMBER: US/10/085,612
; CURRENT FILING DATE: 2002-02-26
; PRIOR APPLICATION NUMBER: 09/144,367
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: 60/271,630
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1345
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-085-612-3

Query Match          67.1%; Score 22.8; DB 14; Length 1345;
Best Local Similarity 92.3%; Pred. No. 7.2;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 9 TGCTACTGGCTGCAGCTGCAGCCCG 34
    |||||
DB 1035 TGCTACTGGCTGCAGCTGCAGCCCTG 1060

RESULT 7
US-09-957-997-1
; Sequence 1, Application US/09957997
; Patent No. US20020150915A1
; GENERAL INFORMATION:
; APPLICANT: Berkenstam, Anders
; APPLICANT: Bertilsson, Gran
; APPLICANT: Blomquist, Patrik
; TITLE OF INVENTION: PROMOTER SEQUENCES
; FILE REFERENCE: 13425-046001
; CURRENT APPLICATION NUMBER: US/09/957,997
; CURRENT FILING DATE: 2001-09-21
; EARLIER APPLICATION NUMBER: SE 0003393-6
; EARLIER FILING DATE: 2000-09-22
; EARLIER APPLICATION NUMBER: 60/238,895
; EARLIER FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 11186
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-957-997-1

Query Match          67.1%; Score 22.8; DB 10; Length 11186;
Best Local Similarity 92.3%; Pred. No. 8.9;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 9 TGCTACTGGCTGCAGCTGCAGCCCG 34
    |||||
DB 11064 TGCTACTGGCTGCAGCTGCAGCCCTG 11089

RESULT 8
US-10-121-960C-14
; Sequence 14, Application US/10121960C
; Publication No. US20030145341A1
; GENERAL INFORMATION:
; APPLICANT: ZHANG, Weisheng
; APPLICANT: CONTAG, Pamela
; APPLICANT: PURCHIO, Anthony
; APPLICANT: HASHIMA, Sandy
; APPLICANT: MA, Shirley
; APPLICANT: NAWOTKA, Kevin
```

```
; TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF MOUSE AND HUMAN
; TITLE OF INVENTION: TRANSCRIPTION CONTROL ELEMENTS ASSOCIATED WITH
; FILE REFERENCE: 9400-0014 / PXE-014.US
; CURRENT APPLICATION NUMBER: US/10/121,960C
; CURRENT FILING DATE: 2002-04-11
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 13035
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: human CYP3A4 gene locus
US-10-121-960C-14

Query Match          67.1%; Score 22.8; DB 12; Length 13035;
Best Local Similarity 92.3%; Pred. No. 9.1;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 9 TGCTACTGGCTGCAGCTGCAGCCCG 34
    |||||
DB 12860 TGCTACTGGCTGCAGCTGCAGCCCTG 12885

RESULT 9
US-10-121-960C-17
; Sequence 17, Application US/10121960C
; Publication No. US20030145341A1
; GENERAL INFORMATION:
; APPLICANT: ZHANG, Weisheng
; APPLICANT: CONTAG, Pamela
; APPLICANT: PURCHIO, Anthony
; APPLICANT: HASHIMA, Sandy
; APPLICANT: MA, Shirley
; APPLICANT: NAWOTKA, Kevin
; TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF MOUSE AND HUMAN
; TITLE OF INVENTION: TRANSCRIPTION CONTROL ELEMENTS ASSOCIATED WITH
; FILE REFERENCE: 9400-0014 / PXE-014.US
; CURRENT APPLICATION NUMBER: US/10/121,960C
; CURRENT FILING DATE: 2002-04-11
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 15185
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: CYP3A4-luc transgene
US-10-121-960C-17

Query Match          67.1%; Score 22.8; DB 12; Length 15185;
Best Local Similarity 92.3%; Pred. No. 9.2;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 9 TGCTACTGGCTGCAGCTGCAGCCCG 34
    |||||
DB 12860 TGCTACTGGCTGCAGCTGCAGCCCTG 12885

RESULT 10
US-10-027-632-72580
; Sequence 72580, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
```

RESULT 12  
US-10-215-432-41/c  
; Sequence 41, Application US/10215432

```

RESULT 14
US-10-027-632-253693
; Sequence 253693, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108627.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006

```

```

QY      10  GCTACTGGCTGCAGCTGCAGCCCG 34
      ||||| ||||| ||||| ||||| ||
Db      285  GCTACAGGCTGCAGCTGCAGCAGCG 309

Search completed: August 27, 2003, 21:14:09
Job time : 179.369 secs

```

```

: PRIOR FILING DATE: 2000-07-12
: PRIOR APPLICATION NUMBER: US 60/198,676
: PRIOR FILING DATE: 2000-04-20
: PRIOR APPLICATION NUMBER: US 60/193,483
: PRIOR FILING DATE: 2000-03-29
: PRIOR APPLICATION NUMBER: US 60/185,218
: PRIOR FILING DATE: 2000-03-24

```

Search completed: August 27, 2003, 21:14:09  
Job time : 179.269 secs

; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24

```

Query Match      59.4%; Score 20.2; DB 13; Length 672;
Best Local Similarity 75.8%; Pred. NO. 77;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

Query Match

Best Local Similarity	75.8%;	Pred. No. 77;	
Matches	25;	Conservative	0; Mismatches 8; Indels 0; Gaps 0;

**OY**            2 GATTCTTTGCTACTGGCTGCAGCTGCAGCCCCG 34  
               ||    ||    ||    ||    ||    ||    ||    ||    ||  
**Dd**            548 GAGCCTGTGCTCCCTGGGGTGAATGCAGCCCAG 580

QY 2 GATCTTTGCTACTGGCTGCAGCTGCAGCCCCG 34  
|| || || || || || || || || || || || || ||  
Db 548 GAGCCTGTGCTCCCTGGGGTGAATGCAGCCCAG 580

```

RESULT 15
US-09-934-455-489
; Sequence 489, Application US/09934455
; Publication No. US20030121070A1
; GENERAL INFORMATION:
; APPLICANT: Adam, Luc
; APPLICANT: Creelman, Robert
; APPLICANT: Dubell, Arnold
; APPLICANT: Heard, Jacqueline
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Kiddie, James
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Reuber, Lynne
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Pineda, Omaira
; TITLE OF INVENTION: Genes for Modifying Plant Traits IV

```

```

/ FILE REFERENCE: MBI-0025
/ CURRENT APPLICATION NUMBER: US/09/934,455
/ CURRENT FILING DATE: 2001-08-22
/ PRIOR APPLICATION NUMBER: 60/227439
/ PRIOR FILING DATE: 2000-08-22
/ PRIOR APPLICATION NUMBER: MBI-0022
/ PRIOR FILING DATE: 2001-11-16
/ PRIOR APPLICATION NUMBER: MBI-0023
/ PRIOR FILING DATE: 2001-04-17
/ NUMBER OF SEQ ID NOS: 516
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 489
/ LENGTH: 994
/ TYPE: DNA
/ ORGANISM: Arabidopsis thaliana
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(711)
/ OTHER INFORMATION: G2547
/ US-09-934-455-489

```

Query Match	59.4%;	Score 20.2;	DB 11;	Length 994;
Best Local Similarity	88.0%;	Pred. No. 80;		
Matches 22; Conservative	0;	Mismatches 3;	Indels 0	

Query Match	59.4%;	Score 20.2;	DB 11;	Length 994;
Best Local Similarity	88.0%;	Pred. No. 80;		
Matches 22; Conservative	0;	Mismatches 3;	Indels 0	